PCT/IL2004/000759
19 AUG 2004

REC'D 03 SEP 2004
WIPO PCT

PA 1202423

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 02, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: ***60/495,898***

FILING DATE: ***August 19, 2003***

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

P. SWAIN
Certifying Officer

BEST AVAILABLE COPY

PATENT APPLICATION SERIAL NO. ______________________

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

08/20/2003 LWONDIM1 00000024 60495898

01 FC:2005 80.00 OP

PTO-1556
(5/87)

*U.S. Government Printing Office: 2002 — 489-267/69033

08/19/03 13281 U.S. PTO

PTO/SB/16 (10-01)
Approved for use through10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No.

INVENTOR(S)

Given Name (first and middle (if any))	Family Name or Surname	Residence (City and either State or Foreign Country)
DANIEL	HARARI	REHOVOT, ISRAEL

22151 U.S. PTO 60/495598 08/19/03

☐ *Additional inventors are being named on the _______ separately numbered sheets attached hereto*

TITLE OF THE INVENTION (500 characters max)

SPLICE VARIANTS OF ErbB LIGANDS, COMPOSITIONS AND USES THEREOF

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

☐ Customer Number [ ] *Type Customer Number here* → *Place Customer Number Bar Code Label here*

OR

☐ Firm or Individual Name	WEBB & ASSOCIATES				
Address	P.O. BOX 2189				
Address					
City	REHOVOT	State		ZIP	78121
Country	ISRAEL	Telephone	972-8-9470830	Fax	972-8-9470860

ENCLOSED APPLICATION PARTS *(check all that apply)*

- ☑ Specification *Number of Pages* 79
- ☑ Drawing(s) *Number of Sheets* 7
- ☐ Application Data Sheet. See 37 CFR 1.76
- ☐ CD(s), Number
- ☑ Other (specify) 68 pages Sequence Listing

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

- ☑ Applicant claims small entity status. See 37 CFR 1.27.
- ☐ A check or money order is enclosed to cover the filing fees
- ☐ The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number:
- ☐ Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($)

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

- ☑ No.
- ☐ Yes, the name of the U.S. Government agency and the Government contract number are: ____________

Respectfully submitted,

SIGNATURE Cynthia Webb

TYPED or PRINTED NAME CYNTHIA WEBB

TELEPHONE

Date 19 Aug 2003

REGISTRATION NO. *(if appropriate)* 39,415

Docket Number: HAR-001

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

SPLICE VARIANTS OF ErbB LIGANDS, COMPOSITIONS AND USES THEREOF

FIELD OF THE INVENTION

The present invention relates to nucleic acid and amino acid sequences of previously unknown ErbB ligands that are splice variants of previously known ErbB ligands and to compositions comprising these sequences, and uses thereof in the diagnosis, treatment, and prevention of diseases and disorders mediated by ErbB receptors.

BACKGROUND OF THE INVENTION

Receptor tyrosine kinases play a key role in the dissemination of cell to cell signaling in organisms typically upon activation via specific activating ligands. Type-1 tyrosine kinase receptors, also known as ErbB/HER proteins, comprise one such receptor tyrosine kinase family, of which the epidermal growth factor receptor (EGFR; ErbB-1) is the prototype. The mammalian/human ErbB family to date consists of four known receptors (ErbB-1 to ErbB-4). Upon ligand binding the receptors dimerize, transducing their signals by subsequent autophosphorylation catalyzed by an intrinsic cytoplasmic tyrosine kinase, and recruiting downstream signaling cascades (reviewed by Yarden and Sliwkowski 2001).

<u>The ErbB ligands</u>

The ErbB receptors are activated by a large number of ligands. This ligand family is encoded in humans by at least eleven independent genes and their splice variants and include the Neuregulins (NRG-1, NRG-2, NRG-3 & NRG-4), the Epidermal Growth Factor (EGF), TGF alpha, Betacellulin, Amphiregulin, Heparin-Binding EGF (HB-EGF), Epiregulin and Epigen (reviewed in Harari et. al. 1999; Harris et. al 2003). These ligands each have a selective repertoire of receptors to which they bind preferentially, each with its own array of differential binding affinities. Typically but not exclusively, the Neuregulins preferably bind to ErbB3 and/or ErbB4, whereas the remaining ligands bind ErbB1. Upon ligand binding, receptor homodimers and heterodimers are typically recruited. ErbB2, which is bound by no known ligand, nevertheless can be actively recruited in a ligand-

dependent manner, as a heterodimer. Depending upon the activating ligand, most homodimeric and heterodimeric ErbB combinations can be stabilized upon ligand binding, thus allowing a complex, diverse downstream signaling network to arise from these four receptors. The choice of dimerization partners for the different ErbB receptors, however, is not arbitrary. Spatial and temporal expression of the different ErbB receptors do not always overlap in vivo, thus narrowing the spectrum of possible receptor combinations that an expressed ligand can activate for a given cell type (reviewed in Harari et al. 1999; Harari and Yarden 2000).

A hierarchical preference for signaling through different ErbB receptor complexes takes place in a ligand-dependent manner. Of these, ErbB-2-containing combinations are often the most potent, exerting prolonged signaling through a number of ligands, likely due to an ErbB-2-mediated deceleration of ligand dissociation. In contrast to possible homodimer formation of ErbB-1 and ErbB-4, for ErbB-2, which has no known direct ligand, and for ErbB-3, which lacks an intrinsic tyrosine kinase activity, homodimers either do not form or are inactive. Heterodimeric ErbB complexes are arguably of importance in vivo. For example, mice defective in genes encoding either NRG-1, or the receptors ErbB-2 or ErbB-4, all result in identical failure of trabeculae formation in the embryonic heart, consistent with the notion that trabeculation requires activation of ErbB-2/ErbB-4 heterodimers by NRG-1 (reviewed in Harari et al. 1999).

The repertoire of ErbB ligands and receptors differs between simpler and more complex organisms. In the worm C. elegans, a single ErbB ligand and receptor are encoded (Moghal and Sternberg 2003). Drosophila melanogaster likewise encodes a single ErbB receptor gene but has an expanded ligand family of four agonists (Vein, Gurken, Spitz and Keren) and a single antagonist, named Argos (Shilo 2003; Table 1). In mammals this has further expanded to genes encoding at least eleven ligands and four receptors. However, no mammalian inhibitory Argos-like ErbB ligand has been described to date. These known ErbB ligands are listed in Table 1.

Table 1:
Agonist and Antagonist Ligands of the ErbB Receptor Tyrosine Kinase Family

	Agonist	Antagonist
C. elegans	Lin-3	
Drosophila	Vein Gurken Spitz Keren	Argos
Mammals	NRG-1 (alpha and beta isoforms) NRG-2 (alpha and beta isoforms) NRG-3 NRG-4 EGF TGF-alpha Betacellulin Amphiregulin Heparin-Binding EGF (HB-EGF) Epiregulin Epigen	

The ErbB ligand receptor-binding EGF domain

Across an evolutionarily diverse selection of organisms, these ligands each harbor an ErbB receptor-binding EGF domains (including the antagonist ligand Argos derived from an invertebrate), which are critical for receptor binding and modulation. Most ligands share the common feature of harboring a single EGF domain and a single transmembrane domain. The EGF domain is found adjacent to the transmembrane domain and on its amino terminal side, thus constituting a component of the ligand ectodomain. The EGF domain is both necessary and sufficient to confer receptor binding and activation. Exceptionally, the Epidermal Growth Factor encodes nine extracellular EGF domains of which only the ninth EGF domain, i.e., that in closest proximity to the transmembrane domain has been shown to confer receptor binding (Carpenter and Cohen 1990). The transmembrane domain tethers the ligand to the cell surface. A complex process of post-translational proteolytic cleavage of the extracellular domain is required to release the tethered EGF domain which in many instances is critical for ligand activation (Harris et al. 2003). However, there do

exist in nature ligands devoid of a transmembrane domain, as is the case for some splice variants of NRG-1 for example. Additionally, a variant of NRG-1 with a truncated EGF domain has been described, albeit reportedly unlikely to be bioactive (Falls 2003).

The ErbB-receptor-binding EGF domains harbor six invariant cysteine residues which are responsible for the formation of three disulfide bridges (considered to form the bridges Cys1-Cys3, Cys2-Cys4 and Cys5-Cys6) these denoted as loops A, B and C (Figure 1 from Harari and Yarden 2000). Besides the conserved cysteines, the receptor-binding EGF domain of these ligands encode numerous conserved and semi-conserved residues, including a Glycine and Arginine residue proximal to Cys-6 (boxed residues in Figure 1 and corresponding to Gly-40 & Arg-42 or Gly-39 & Arg-41 for synthetic peptides encoding the ligand-binding EGF domain of TGF-alpha and epidermal growth factor respectively as defined by others (Jorissen et al. 2003)). The conservation of these Glycine and Arginine residues are not coincidental. Substitutional mutagenesis of these residues severely compromises ligand binding or function (Campion and Niyogi 1994; Groenen et al. 1994; Summerfield et al. 1996).

Drosophila Melanogaster Argos

The inhibitory Drosophila melanogaster ligand Argos harbors an EGF domain which harbors a B-loop which is larger than that for the activatory ligands (Figure 1). Despite this divergence from the remainder of the ErbB ligand family, the Argos EGF domain binds directly to the Drosophila EGF Receptor (Jin et al. 2000; Vinos and Freeman 2000). The Argos EGF domain plays an essential role not just in receptor binding, but also in the ligand's antagonist function. A domain swap of the Argos EGF domain into the agonist ligand Vein, converted this activatory ligand into an inhibitor (Schnepp et al. 1998). Furthermore, Argos blocks the binding of secreted Spitz to the Drosophila EGF receptor, suggestive that the inhibitory ligand competitively displaces agonist ligand binding (Jin et al. 2000).

In the C-loop, Drosophila melanogaster Argos harbors the canonical Glycine and Arginine residues typical for this ligand family (Boxed region; Figure 1; equivalent to Gly39 and Arg41 of EGF (Groenen et al. 1994)). However, this otherwise invariant Arginine residue has been substituted to a Histidine, in Argos sequenced from Musca domestica, another insect species, demonstrating that absolute

conservation at this residue is not required for Argos function (Howes et al. 1998). This finding has been re-represented in Figure 2, as a multiple alignment for three insect species. The significance of the Arg to His substitution in Musca domestica Argos should not be underestimated. A panel of substitution mutations of EGF Arg41 (or the corresponding Arg42 of TGF alpha) were shown to decrease ligand-binding affinity by more than 100-fold (Campion and Niyogi 1994; Defeo-Jones et al. 1989; Engler et al. 1992).

From these combined data, it may be construed that the C-loop of Argos cannot be considered responsible (or at least entirely responsible) for Argos inhibitory function. In support of this hypothesis, the replacement of the Argos C-loop with that from the stimulatory Drosophila ligand Spitz, results in the formation of a chimeric protein that retains moderate inhibitory activity (Howes et al. 1998).

ErbB ligands have been shown to be essential in induction and propagation of cell proliferation and are involved in many cell-signaling pathways in a wide variety of normal and malignant physiological events. Therefore, both agonists and antagonists of the ErbB signaling pathways have enormous therapeutic potential (reviewed by Mendelsohn and Baselga, 2003).

The above described ErbB ligands and methods of using same emphasize the phenomenon that different ErbB ligands may have different structure and function. Novel splice variants of ErbB ligands are likely to have a physiological role, whether systemic or tissue specific.

Therefore, there is a recognized need for, and it would be highly advantageous to isolate and characterize ErbB ligand splice variants, that may include truncations, deletions, alternative exon splicing or translatable intronic sequences, which alter the composition or length of the receptor-binding EGF domain.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide novel ErbB ligand splice variants, including truncation variants, deletion variants, alternative exon usage, and intronic sequences, that all comprise at least one component of the EGF domain responsible for receptor binding. The invention relates to isolated polynucleotides encoding these novel variants of ErbB ligands, including recombinant DNA constructs comprising these polynucleotides, vectors comprising the constructs, host

cells transformed therewith, and antibodies that specifically recognize one or more epitope present on such splice variants.

It is another object of the present invention to provide vectors, including expression vectors containing the polynucleotides of the invention, cells engineered to contain the polynucleotides of the present invention, cells genetically engineered to express the polynucelotides of the present invention, and methods of using same for producing recombinant ErbB ligand splice variants according to the present invention.

It is a further object of the present invention to provide synthetic peptides comprising the novel amino acid sequences disclosed herein. It is explicitly to be understood that the novel splice variants disclosed herein as ErbB ligands, whether deduced from conserved genomic DNA sequences, deduced from cDNA sequences, or derived from other sources, may be produced by any suitable method involving recombinant technologies, synthetic peptide chemistry or any combination thereof.

It is a yet another object of the present invention to provide pharmaceutical compositions comprising the novel ErbB ligand splice variant or polynucleotide encoding same. It is yet further object of the present invention to provide methods for the diagnosis and treatment of ErbB receptor related diseases comprising administering to a subject in need thereof a pharmaceutical composition comprising as an active ingredient a novel ErbB ligand or a polynucleotide encoding same.

According to one aspect, the present invention provides ErbB ligand splice variant polypeptides and polynucleotides encoding same. Novel isoforms and putative isoforms of known ErbB ligands are disclosed, that are characterized in that they do not comprise the C-loop of the receptor-binding EGF domain. In other words, the unifying feature of the splice variants of the present invention is that they lack cysteines 5 and 6 of the invariant six cysteines of hitherto known ErbB ligand receptor-binding EGF domains.

According to one embodiment, the present invention provides novel mature polypeptides having ErbB receptor agonist or antagonist activity, as well as fragments, analogs and derivatives thereof. According to some embodiments, the polypeptides of the present invention are of non-mammalian vertebrate origin. According to other embodiments, the polypeptides of the present invention are of mammalian origin. According to other embodiments the polypeptides are of human origin.

According to a one embodiment the present invention provides a polypeptide comprising a splice variant of an ErbB ligand encoded by differential exon usage comprising a truncated ErbB-Receptor-binding EGF domain devoid of the C-loop of the EGF domain.

According to some embodiments the present invention provides ErbB ligand splice variants, comprising the sequence of SEQ ID NOs: 73 to 84 and 93 to 127.

According to another embodiment the present invention provides polynucleotides encoding for the ErbB ligand splice variants, including an isolated polynucleotide encoding a polypeptide comprising the sequence of SEQ ID NOS: 128 to 139 and 148 to 192.

It is to be understood that the present invention encompasses all active fragments, variants and analogs of the sequences disclosed herein that retain the biological activity of the sequence from which they are derived, with the proviso that said variants and analogs are devoid of the C-loop of the EGF domain.

The invention also provides a polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide encoding the amino acid sequence of SEQ ID NOS:73 to 84 and SEQ ID NOS:93 to 127, or fragments of said polynucleotide sequences. The invention further provides a polynucleotide sequence comprising the complement of the polynucleotide sequence encoding the amino acid sequence of SEQ ID NOS:73 to 84 and SEQ ID NOS:93 to 127, or fragments or variants of said polynucleotide sequence.

According to some embodiments, the isolated polynucleotides of the present invention include a polynucleotide comprising the nucleotide sequence of SEQ ID NOS:128 to 139 and SEQ ID NOS:148 to 192, or fragments, variants and analogs thereof. The present invention further provides the complement sequence for a polynucleotide having SEQ ID NO:128 to 139 and SEQ ID NOS:148 to 192 or fragments, variants and analogs thereof. The polynucleotide of the present invention also includes a polynucleotide that hybridizes to the complement of the nucleotide sequence of SEQ ID NOS:128 to 139 and SEQ ID NOS:148 to 192 under stringent hybridization conditions.

According to yet another embodiment, the present invention provides an expression vector containing at least a fragment of any of the polynucleotide sequences disclosed. In yet another embodiment, the expression vector containing the

polynucleotide sequence is contained within a host cell. The present invention further provides a method for producing the polypeptides according to the present invention comprising a) culturing the host cell containing an expression vector containing at least a fragment of the polynucleotide sequence encoding an ErbB ligand splice variant under conditions suitable for the expression of the polypeptide; and b) recovering the polypeptide from the host cell culture.

According to another aspect the present invention also provides a method for detecting a polynucleotide which encodes an ErbB variant ligand in a biological sample comprising the steps of: a) hybridizing the complement of the polynucleotide sequence which encodes SEQ ID NOS:73 to 84 and SEQ ID NOS:93 to 127 to nucleic acid material of a biological sample, thereby forming a hybridization complex; and b) detecting the hybridization complex, wherein the presence of the complex correlates with the presence of a polynucleotide encoding an ErbB variant ligand in the biological sample. According to one embodiment the nucleic acid material of the biological sample is amplified by the polymerase chain reaction prior to hybridization.

According to yet another aspect the present invention provides a pharmaceutical composition comprising a polypeptide having the amino acid sequence of SEQ ID NOS:73 to 84 and SEQ ID NOS:93 to 127 or a polynucleotide encoding same, further comprising a pharmaceutically acceptable diluent or carrier.

According to further aspects the present invention provides a purified inhibitor or antagonist of the ErbB ligand splice variant of the present invention. The inhibitor or antagonist may be selected from the group consisting of antibodies, peptides, peptidomimetics and small organic molecules. The inhibitor, preferably a specific antibody, has a number of applications, including identification, purification and detection of variant ErbB ligand, specifically any antibody capable of recognizing an epitope present on the ErbB ligand splice variant devoid of the C-loop of the EGF domain, that is absent form the known counterparts that include the C-loop of the EGF receptor binding domain.

According to one embodiment, the present invention provides a purified antibody which binds to at least one epitope of a polypeptide comprising the amino acid sequence of SEQ ID NOS:73 to 84 and 93 to 127, or specific fragments, analogs

and variants thereof, with the proviso that the epitope is absent on the known counterpart ErbB ligands.

Further aspects of the present invention provide methods for preventing, treating or ameliorating an ErbB receptor related disease or disorder, comprising administering to a subject in need thereof a pharmaceutical composition comprising as an active ingredient an ErbB ligand splice variant, as disclosed hereinabove.

According to one embodiment, the present invention provides a method for preventing, treating or ameliorating an ErbB receptor related disease or disorder, comprising administering to a subject in need thereof a pharmaceutical composition comprising as an active ingredient a polypeptide comprising the SEQ ID NOS:73 to 84 and 93 to 127.

According to another embodiment, the present invention provides a method for preventing, treating or ameliorating an ErbB receptor related disease or disorder, comprising administering to a subject in need thereof a pharmaceutical composition comprising as an active ingredient a polynucleotide encoding a polypeptide comprising the SEQ ID NOS:73 to 84 and 93 to 127.

According to another embodiment, the present invention provides a method for preventing, treating or ameliorating an ErbB receptor related diseases or disorder, comprising administering to a subject in need thereof a pharmaceutical composition comprising as an active ingredient a polynucleotide comprising the SEQ ID NOS:128 to 139 and 148 to 192.

According to yet another embodiment, the ErbB receptor related diseases or disorders are selected from the group consisting of neoplastic disease, hyperproliferative disorders, angiogenesis, restenosis, wound healing, psychiatric disorders, neurological disorders and neural injury.

As it is anticipated that at least some of the novel ErbB splice variants having a truncated EGF receptor-binding domain lacking the C-loop of the intact EGF domain, may act as antagonists rather than agonists it is to be understood that these variants will be useful to prevent or diminish any pathological response mediated by a ligand agonist. Thus, the neoplastic, hyperproliferative, angiogenic or other response may be attenuated or even abrogated by exposure or treatment with an antagonist according to the present invention.

Furthermore, if an agonist ligand predisposes stem cells to proliferate, survive,

migrate, enter or commit to a specific lineage, then exposure or treatment with an antagonist would have the potential to alter the lineage commitment or differentiation pattern, or enhance proliferation prior to commitment to a given cell lineage.

According to yet further aspects the present invention provides methods for selectively enhancing or promoting the proliferation or differentiation of stem cells expressing ErbB receptors, comprising exposing the stem cells to an ErbB ligand splice variant, according to the present invention. Preferably, said stem cells are of neural, cardiac or pancreatic lineages, as ErbB ligands are known in the art to be involved in the development of these lineages.

According to one embodiment, the present invention provides a method for selectively enhancing or promoting the proliferation or differentiation of stem cells expressing ErbB receptors, comprising exposing the stem cells to an ErbB ligand splice variant comprising the SEQ ID NOS:73 to 84 and 93 to 127.

More preferably said stem cells are selected from neural, cardiac or pancreatic stem cell lineages.

According to further aspects the present invention provides methods of inhibiting the expression of the ErbB ligand splice variant by targeting the expressed transcript of such splice variant using antisense hybridization, siRNA inhibition and ribozyme targeting.

The present invention is explained in greater detail in the description, figures and claims below.

BRIEF DESCRIPTION OF THE FIGURES

FIGURE 1

Multiple sequence alignment of the evolutionarily conserved receptor-binding EGF domains for different known ErbB-ligands identified forin worms (C. elegans), insects (Drosophila melanogaster) and mammals (humans or mice). Sequences shaded in grey demonstrate invariant residues in this alignment. Six cysteine residues are thought to be required for the formation of three disulfide loops within the domain for all these known ligands. Additionally, an invariant Glycine and Arginine residue is also considered critical for ligand-receptor binding (boxed region). This multiple alignment was generated by ClustalX (version 1.81) using the following protocol: The mammalian sequences were independently aligned by ClultalX (default parameters).

This was repeated for the invertebrate ligands. These alignments were then treated as independent profiles, where the profile of mammalian sequences was aligned against the profile of invertebrate sequences, once again using clustalX (profile mode). All calculations were performed using default program parameters.

FIGURE 2

Multiple alignment of Argos primary protein sequences published for three independent insect species, Drosophila melanogaster, Drosophila virilis and Musca domestica. Two cysteine-rich domains defined as A1 and A2 and the EGF domain are marked in bold-set and underlined. The definitions demarking these domains have been borrowed from elsewhere (Howes et. al, 1999). Regions of highly conserved residues indicate the presence of critical domains within the Argos protein sequences. Similarly, the Musca domestica protein seequence demonstrates that an invariant Arg residue found for all other known receptor agonists (see Figure 1) is not necessarily conserved in insect Argos (boxed region).

* = Invariant residues, : = Conserved residues, . = Semi-conserved residues.

FIGURE 3

Shows multiple sequence alignment of the receptor-binding EGF domain encoded by different mammalian ErbB-ligands. Multiple sequence alignment of the receptor-binding EGF domain encoded by different mammalian ErbB-ligands were used as an input from which to generate a sequence profile in order to perform profile searches against various databases using a Compugen Biocellerator. This alignment was generated by ClustalX version 1.81 and with minor manual modification. * = Invariant residues, : = Conserved residues, . = Semi-conserved residues.

FIGURE 4

Examination of the genomic locus encoding "Exon A" of the EGF domain for the Neuregulin/EGF ligand family. The genomic sequence encoding Exon A for each ligand was extracted from the NCBI human (or where indicated mouse) genomic database. The genomic sequence was then translated, this including extended sequence running into and beyond the 5' exon:intron splice junction which typically

demarks the end of Exon A. This 'extended Exon A' potentially encodes an invariant in-frame stop codon positioned at precisely the same coordinate for all ErbB ligands relative to cysteine 4 of the EGF domain. The protein sequences of the full-length EGF domains are aligned in this figure against the translated sequence of extended Exon A. Exon A and Exon B are alternatively shaded. The dotted lines (....) indicate that the exon-encoding sequences extend beyond this alignment. The protein sequences present in this figure are listed in this patent as indicated (SEQ ID NOS:14-26, and 73-84). The nucleotide sequences encoding extended Exon A for each ligand are also provided (SEQ ID NOS:128-139). Human and mouse variant Epigen sequences are provided and serve to exemplify that the "Extended Exon A" topology is conserved not only for different ligands within a single species, but is also conserved for different mammalian species. The human Epigen variant EGF domain sequence provided in this figure, which is truncated after the conserved fourth encoded cysteine of the domain was predicted from genomic data by its similarity to the mouse Epigen protein sequence (tblastn search; performed using the NCBI server). A similar genomic topology was found for genes encoding other mouse ErbB ligands.

FIGURE 5

EGF domains other than the mammalian ErbB-binding domains are encoded at the genomic level in a heterogeneous manner.

FIGURE 5A shows a schematic diagram of the EGF domain structure for TGF alpha, EGF and Notch-1. The proteins TGF alpha, EGF and Notch-1 harbor one, nine and thirty-six EGF domains within their respective sequences as shown (diagram is not to scale). EGF domains are represented as boxes. The transmembrane domain of both EGF and TGF-alpha are represented as vertical black bars. Other unrelated domains are ignored in this diagram. The EGF domains responsible for receptor binding (for both EGF and TGF alpha) are denoted as shaded boxes followed by a star (*). Epidermal Growth Factor harbors an additional eight EGF domains not thought to directly bind receptor. Notch-1 is not considered an ErbB ligand and is shown here as an example of an unrelated protein which also harbors EGF domains (unshaded boxes).

FIGURE 5B Examination of the genomic locus encoding different EGF domains for human TGF alpha, EGF and Notch-1. The protein sequences for TGF alpha (i), EGF (ii) and Notch-1 (iii) were blasted against the human genomic database (tblastn), to examine the exon structure for these genes. The EGF domains of these protein sequences were identified using the SMART database with manual adjustment, where flanking sequences have been ignored. These domain sequences were aligned (Clustalx version 1.81; standard parameters). Dark and light shading indicate the genomic topology demarking exon-exon boundaries within a particular EGF domain. The coordinates of each EGF domain is given in each case. For example, the first EGF domain which spans amino acids 24-57 for Notch-1 is shown as EGF_24_57. The protein sequences and genomic sequences used to examine TGF alpha, EGF and Notch-1 were derived from the NCBI accessions [P01135, NT_022184.9], [NP_001954.1, NT_028147.9] and [AAG33848, NT_024000.13] respectfully. Of the aligned domains, the exceptional examples of ErbB-recepor-binding EGF domains are typed in bold-set and demarked with a star (*). Of the forty four EGF domains examined which do not directly bind ErbB receptors (thirty six domains for Notch-1 and eight domains for EGF), only two of these (Notch-1 EGF domains number 1 and 30) harbor an exon-exon boundary which splits Cysteine 1-4 and Cys 5-6. The first EGF domain of Notch-1 is not fully shaded, due to the lack of this segment of genomic sequence found in the BLAST alignment.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to (i) novel ErbB ligand isoforms identified as splice variants of at least one known ErbB ligand; (ii) polynucleotide sequences encoding the novel splice variants; (iii) oligonucleotides and oligonucleotide analogs derived from said polynucleotide sequences; (v) antibodies recognizing said splice variants; (vi) peptides or peptide analogs derived from said splice variants; and (vii) pharmaceutical compositions; and (viii) methods of employing said polypeptides, peptides or peptide analogs, said oligonucleotides and oligonucleotide analogs, and/or said polynucleotide sequences to regulate at least one ErbB receptor mediated activity.

While conceiving the present invention it was hypothesized that additional,

previously unknown, ErbB ligands may exist. Splice variants, which occur in over 50% of human genes, are usually overlooked in attempts to identify differentially expressed genes, as their unique sequence features including donor-acceptor concatenation, an alternative exon, an exon and a retained intron, complicate their identification. However, splice variants may have an important impact on the understanding of disease development and may serve as valuable markers in various pathologies.

ErbB Ligand Splice Variants

The exact definition of what may constitute the boundaries of an ErbB-ligand receptor binding EGF domain is a matter of dispute. A conservative and limiting view is that it spans Cysteine 1 to Cysteine 6 (C1–C6) precisely (e.g. Howes et al. 1998). Even smaller sub-domains of this region were reported to weakly bind to receptors and to induce very low levels of biological activity (reviewed in Groenen et al. 1994). An alternative definition is based upon the natural cleavage pattern of pro-ligands, in which EGF-domain harboring peptides of varying length are generated after proximal and distal cleavage events (Harris et al. 2003). Yet other definitions rely upon biochemical and bioactivity analyses of synthetic and recombinant peptides of varying length, to reconstitute "typical" ligand function. From such analyses, it is apparent that additional carboxy and amino terminal sequences flanking C1-C6 are required to reconstitute ligand function. The exact length required for "typical" function may differ from ligand to ligand, as has been experimentally demonstrated in studies based upon binding and bioactivity assays (Barbacci et al. 1995; Groenen et al. 1994; Jones et al. 1999). Even so, it is evident that such definitions may vary depending on the biological assay performed. For example, biological assays based upon elucidation of binding affinity for a synthetic ligand peptide alone may demonstrate that a particular ligand of defined length binds very weakly. However, potent mitogenic low affinity ligands have been described in nature (for example Tzahar et al. 1998). Thus a disparity exists between these two biological parameters.

Although each Neuregulin encodes only a single EGF domain, both NRG-1 and NRG-2 harbor splice variants in which the carboxy-terminus of the EGF domain can be encoded by two alternative exons (the resultant variants termed alpha and beta). These alternatively encoded ligands harbor different binding affinities and capacities to heterodimerize with the four different ErbB receptors (reviewed by Falls, 2003).

The ability to generate alpha and beta isoforms for NRG1 and NRG2 are reflected at the genomic level, where the carboxyl terminus of the EGF domain is encoded by alternate exons. More specifically, for both NRG1 and NRG2, a single exon encodes the amino-terminal component of the EGF domain, spanning C1-C4 and constituting the A-loop and B-loop of the EGF domain. An alternative choice of exons encode the remainder of the domain, which harbors C5-C6; the C-loop of the EGF domain (Crovello et al. 1998). Interestingly, all other members of the ErbB ligand family also share a similar segmented exon domain structure, precisely encoding C1-C4 and C5-C6 of the receptor-binding EGF domains on adjacent exons. However, for all these ligands other than NRG1 and NRG2, there has been no evidence to indicate that they encode alpha and beta alternative isoforms of the EGF domain, thus the evolutionary forces which are maintaining these conserved exon-exon topologies at the genomic level remains enigmatic (Harris et al. 2003; Additionally disclosed by D. Harari, BigRock Seminar, the Weizmann Institute of Science, February 5th, 2001). The functional significance of the maintenance of this exon-exon structure of the receptor-binding EGF domains has remained unresolved, and is the major focus of the present invention.

To date only one ErbB ligand having antagonist activity has been identified, namely the Argos ligand from different insects. One major objective of the present invention is to identify additional ErbB ligands that may possess inhibitory activity, especially naturally occurring ligands, preferably from vertebrate species, more preferably from mammalian species, most preferably from humans. Besides the importance of the EGF domain, Drosophila Argos harbors two additional cysteine rich regions, which have been defined as A1 and A2 (Howes et al. 1998). The multiple alignment of Argos from three species demonstrates that as for the EGF domain, domains A1 and A2 and adjacent sequences are highly conserved (Figure 2), supporting an important physiological function of these domains in the function of the protein. This multiple alignment also demonstrates conservation of sequence for the EGF domain and flanking carboxyl-terminal sequence (Figure 2).

Before describing the present proteins, nucleotide sequences, the compositions comprising same and methods of use thereof, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents

described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" includes a plurality of such host cells, reference to the "antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies, which are reported in the publications which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

Definitions

ErbB ligand, as used herein, refers to the amino acid sequences of substantially purified ErbB ligand obtained from any species, particularly higher vertebrates, especially mammalian, including bovine, ovine, porcine, murine, equine, and preferably human, from any source whether natural, synthetic, semi-synthetic, or recombinant.

As used herein in the specification and in the claims section that follows, the phrase "complementary polynucleotide sequence" includes sequences which originally result from reverse transcription of messenger RNA using a reverse transcriptase or any other RNA dependent DNA polymerase. Such sequences can be subsequently amplified in vivo or in vitro using a DNA dependent DNA polymerase.

As used herein in the specification and in the claims section that follows, the

phrase "genomic polynucleotide sequence" includes sequences which originally derive from a chromosome and reflect a contiguous portion of a chromosome.

As used herein in the specification and in the claims section that follows, the phrase "composite polynucleotide sequence" includes sequences which are at least partially complementary and at least partially genomic. A composite sequence can include some exonal sequences required to encode a polypeptide, as well as some intronic sequences interposing therebetween. The intronic sequences can be of any source, including of other genes, and typically will include conserved splicing signal sequences. Such intronic sequences may further include cis acting expression regulatory elements.

As used herein in the specification and in the claims the phrase "splice variants" refers to naturally occurring nucleic acid sequences and proteins encoded therefrom which are products of alternative splicing. Alternative splicing refers to intron inclusion, exon exclusion, alternative exon usage or any addition or deletion of terminal sequences, which results in sequence dissimilarities between the splice variant sequence and other wild-type sequence(s). Although most alternatively spliced variants result from alternative exon usage, some result from the retention of introns not spliced-out in the intermediate stage of RNA transcript processing.

An "allele" or "allelic sequence", as used herein, is an alternative form of the gene encoding an ErbB ligand. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding an ErbB ligand as used herein include those with deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent ErbB ligand. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding a particular ErbB ligand, and improper or unexpected hybridization to alleles, with a locus other than the normal chromosomal locus for the polynucleotide

sequence encoding the ErbB ligand. The encoded protein may also be "altered" and contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent ErbB ligand. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological or immunological activity of the ErbB ligand is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine, glycine and alanine, asparagine and glutamine, serine and threonine, and phenylalanine and tyrosine.

"Amino acid sequence", as used herein, refers to an oligopeptide, peptide, polypeptide, or protein sequence, and fragment thereof, and to naturally occurring or synthetic molecules. Fragments of ErbB ligands are preferably about twenty to about forty amino acids in length and retain the biological activity or the immunological activity of the intact ligand. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, amino acid sequence, and like terms, are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

"Amplification" as used herein refers to the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) technologies well known in the art (Dieffenbach, C. W. and G. S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y.).

The term “activatory ligand” or “agonist”, as used herein, refer to a ligand which upon binding stimulates ErbB signaling in a receptor-dependent manner. The term “inhibitory ligand” or “antagonist”, as used herein, refers to a ligand which in the short term and/or longer term inhibits ErbB signaling in a receptor-dependent manner. Without contradiction, under certain circumstances, a ligand may be correctly described either as activatory and inhibitory, depending on the environmental and experimental context in which it has been described.

The term “inhibitory ligand” or "antagonist", as used herein interchangeably, refers to a molecule which, when bound to an ErbB receptor, decreases the amount or

the duration of the effect of the biological or immunological activity of a known ligand of that receptor. Antagonists may include proteins, peptides, nucleic acids, antibodies or any other molecules which decrease the effect of a known ErbB ligand.

As used herein, the term "antibody" refers to intact molecules as well as fragments thereof, such as Fab, $F(ab')_2$, and Fv, which are capable of binding the epitopic determinant. Antibodies that bind ErbB ligand polypeptides can be prepared using intact polypeptides or fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal can be derived from the translation of RNA or synthesized chemically and can be conjugated to a carrier protein, if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin and thyroglobulin, keyhole limpet hemocyanin. The coupled peptide is then used to immunize the animal (e.g., a mouse, a rat, or a rabbit).

The term "antigenic determinant", as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as antigenic determinants. An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense", as used herein, refers to any composition containing nucleotide sequences which are complementary to a specific DNA or RNA sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules include peptide nucleic acids and may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and block either transcription or translation. The designation "negative" is sometimes used in reference to the antisense strand, and "positive" is sometimes used in reference to the sense strand.

The term "biologically active", as used herein, refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural,

recombinant, or synthetic ErbB ligand, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term “Active fragment” refers to any variant with the truncated domain lacking the C-loop as the minimal receptor binding fragment. An active fragment may be defined as any fragment having less than the six conserved cysteines of the intact EGF domain capable of binding to at least one ErbB receptor subtype. Preferably the term active fragment refers to any fragment having less than the six conserved cysteines of the intact EGF domain capable of binding to at least one ErbB receptor subtype, further comprising flanking amino acid sequences known to increase the receptor binding and/or ligand induced receptor mediated activity.

The terms "complementary" or "complementarity", as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, the sequence "A--G--T" binds to the complementary sequence "T--C--A". Complementarity between two single-stranded molecules may be "partial", in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" as used herein refers broadly to any composition containing the given polynucleotide sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding a novel ErbB ligand splice variant according to the present invention, or specific fragments thereof may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS) and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

A "deletion", as used herein, refers to a change in the amino acid or nucleotide sequence and results in the absence of one or more amino acid residues or nucleotides.

The term "derivative", as used herein, refers to the chemical modification of a nucleic acid encoding or complementary to an ErbB ligand or to the chemical modification of the encoded ErbB ligand. Such modifications include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative encodes a polypeptide which retains the biological or immunological function of the natural molecule. A derivative polypeptide is one which is modified by glycosylation, pegylation, or any similar process which retains the biological or immunological function of the polypeptide from which it was derived.

The term "homology", as used herein, refers to a degree of sequence similarity in terms of shared amino acid or nucleotide sequences. There may be partial homology or complete homology (i.e., identity). For amino acid sequence homology amino acid similarity matrices may be used as are known in different bioinformatics programs (e.g. BLAST, FASTA, Smith Waterman). Different results may be obtained when performing a particular search with a different matrix. Degrees of homology for nucleotide sequences are based upon identity matches with penalties made for gaps or insertions required to optimize the alignment, as is well known in the art.

The term "humanized antibody", as used herein, refers to antibody molecules in which amino acids have been replaced in the non-antigen binding regions in order to more closely resemble a human antibody, while still retaining the original binding ability.

The term "hybridization", as used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

An "insertion" or "addition", as used herein, refers to a change in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, as compared to the naturally occurring molecule.

"Microarray" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The term "modulate", as used herein, refers to a change in the activity of at least one ErbB receptor mediated activity. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional or immunological properties of an ErbB ligand.

"Nucleic acid sequence" as used herein refers to an oligonucleotide, nucleotide,

or polynucleotide, and fragments thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand. "Fragments" are those nucleic acid sequences which are greater than 60 nucleotides than in length, and most preferably includes fragments that are at least 100 nucleotides in length.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20 to 25 nucleotides, which can be used in PCR amplification or a hybridization assay, or a microarray. As used herein, oligonucleotide is substantially equivalent to the terms "amplimers", "primers", "oligomers", and "probes", as commonly defined in the art.

The term “peptide nucleic acid" (PNA) as used herein refers to nucleic acid “mimics”; the molecule’s natural backbone is replaced by a pseudopeptide backbone and only the four-nucleotide bases are retained. The peptide backbone ends in lysine, which confers solubility to the composition. PNAs may be pegylated to extend their lifespan in the cell where they preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P. E. et al. (1993) Anticancer Drug Des. 8:53-63).

The term "portion", as used herein, with regard to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from five amino acid residues to the entire amino acid sequence minus one amino acid. Thus, a protein "comprising at least a portion of the amino acid sequence of SEQ ID NO:1" encompasses the full-length PNIN and fragments thereof.

The term "sample", as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acid encoding an ErbB ligand, or fragments thereof, or the encoded polypeptide itself may comprise a bodily fluid, extract from a cell, chromosome, organelle, or membrane isolated from a cell, a cell, genomic DNA, RNA, or cDNA in solution or bound to a solid support, a tissue, a tissue print, and the like.

The terms "specific binding" or "specifically binding", as used herein, refers to that interaction between a protein or peptide and an agonist, an antibody and an antagonist. The interaction is dependent upon the presence of a particular structure (i.e., the antigenic determinant or epitope) of the protein recognized by the binding

molecule. For example, if an antibody is specific for epitope "A", the presence of a protein containing epitope A (or free, unlabeled A) in a reaction containing labeled "A" and the antibody will reduce the amount of labeled A bound to the antibody.

The terms "stringent conditions" or "stringency", as used herein, refer to the conditions for hybridization as defined by the nucleic acid, salt, and temperature. These conditions are well known in the art and may be altered in order to identify or detect identical or related polynucleotide sequences. Numerous equivalent conditions comprising either low or high stringency depend on factors such as the length and nature of the sequence (DNA, RNA, base composition), nature of the target (DNA, RNA, base composition), milieu (in solution or immobilized on a solid substrate), concentration of salts and other components (e.g., formamide, dextran sulfate and/or polyethylene glycol), and temperature of the reactions (within a range from about 5°C below the melting temperature of the probe to about 20°C to 25°C below the melting temperature). One or more factors be may be varied to generate conditions of either low or high stringency different from, but equivalent to, the above listed conditions.

The term "substantially purified", as used herein, refers to nucleic or amino acid sequences that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated.

A "substitution", as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation", as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. Such "transformed" cells include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome. They also include cells which transiently express the inserted DNA or RNA for limited periods of time.

A "variant" of an ErbB ligand, as used herein, refers to an amino acid sequence

that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant may have "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

The search for novel inhibitory ligands by a bioinformatics approach:

Utilizing a methodology of sequence comparison, it has been possible to identify homologous ErbB ligand agonists by a bioinformatics approach (e.g. (Harari et al. 1999)). However, despite the wealth of sequence data that is publicly available, no naturally known mammalian inhibitory ErbB ligand has been described in the literature to date. Indeed a preliminary BLAST-based database search failed to identify mammalian genes with sequences sufficiently similar that of insect Argos-like proteins to be readily identified (data not shown).

Thus, it was decided to perform searches for sequences that may harbor EGF-like domains with a profile somewhat typical to that already known for members of the mammalian ErbB-ligand family. It should be noted that this search is biased to the identification of ligand agonists, as all known mammalian ligands to date are agonists. However, if the EGF domain of mammalian ErbB antagonist ligands are sufficiently similar to that of their agonist counterparts, it may be possible to identify them by sequence similarity search. Protein sequences for different mammalian ligands were therefore retrieved from the NCBI server (see Materials and Methods & Tables 4 and 5). Approximate identification of the coordinates in which the receptor-binding EGF domains for each ligand was revealed and defined by the SMART server. These domains were arbitrarily lengthened to provide a greater span of amino and carboxyl sequences which may be helpful for the identification of novel ligands, and were subsequently aligned using ClustalX. Minor modification to the sequence alignments were performed manually (Figure 3).

This multiple sequence alignment was subsequently used to create a profile using the program PROFILEWEIGHT (see materials and methods). Translated profile searches were then performed against the EST databases provided at the EMBL site (see materials and methods). At the time of these searches, the EST database was split into five partitions at the EMBL site and each partition was independently scanned by TPROFILESEARH. These searches were performed using global alignments and the choice of gap opening penalties and gap extension penalties (GOP & GEP) being set at (10 & 1) or (12 & 1) respectively with a predefined output of 500 sequences to be aligned per search . No novel ESTs with an obvious encoded sequence profile similar to that typical to the EGF domain of ErbB ligands were identified.

Since it has already been observed that the exon organization encoding all mammalian ErbB ligands at the site of the EGF domain is conserved, it was decided to explore the possibility that alternative ErbB splice variants encoding partial, alternative or truncated EGF domains may be expressed. For example, a truncated form of NRG1, encoding a partial EGF domain up to cysteine 4, followed by a stop codon has been reported (Falls 2003). Splice isoforms can be better characterized when the variants are examined in the context of the genomic sequence encoding each gene.

It was thus decided to extract co-currently the genomic sequences encoding the mammalian ErbB ligands. As a matter of convenience, nomenclature is provided herein to better describe the exons that typically encode the receptor-binding EGF domain for the mammalian ErbB ligands. The first exon encoding the first component of the receptor-binding EGF domain of ErbB ligands (including C1-C4) is described herein as "Exon A" of the EGF domain. The second exon encoding the second component of the EGF domain (including C5-C6) is described herein as "Exon B" of the EGF domain. In the case of NRG1 and NRG2. which harbor alternative (alpha and beta) carboxyl isoforms of the EGF domain, these are considered herein as exon B (for alpha isoforms) or exon B' (for beta isoforms) of the EGF domain. Genomic sequences encoding the different mammalian ErbB ligands were extracted from the NCBI database (See Materials and Methods and Tables 4 and 5). For each gene, the genomic region encoding Exon A including flanking sequences, was identified and translated (using Transeq). A surprising result was observed. Not only is the position

of the exon-exon junction for Exon A and Exon B conserved for all mammalian ErbB ligands, in what would typically be considered as "intronic" region just beyond Exon A, an invariant stop codon has been found and is encoded both in-frame and immediately downstream of Exon A (Figure 4). This provides indirect evidence to support that alternative isoforms of all mammalian ligands may exist in which the encoded proteins harbor truncated EGF domains. Specifically, such splice variants would encode the EGF domain to one amino acid beyond Cysteine 4 (Figure 4) as a result of the extension in length of exon A of the EGF domain.

An examination of the expanded exon A nucleotide sequence (sequence ID #170-181) demonstrates that for each ligand a common consensus pattern leading to the termination of the translation product. The sequences harbor the consensus G,TXX, where the comma denotes the codon reading frame and TXX encodes a stop codon. The di-nucleotide motif "GT" is required to maintain the evolutionarily conserved exon:intron splice junction that is observed at this site (Darnell et. al. 1986).

Thus, an initial hypothesis is provided that the evolutionarily conserved genomic topology of the EGF domain is preserved in order to allow the generation of ErbB-ligand splice variants which are truncated after cysteine-4 of the EGF domain. A negative hypothesis to this concept, is that the exon-exon structure encoding the mammalian ErbB ligand receptor-binding EGF domains has nothing to do with the formation of splice variants, but rather is a result of the general genomic topology found for EGF domain sequences (for reasons that may be known or unknown). EGF domains are commonly encoded by many proteins, with functions that in the most part are unrelated to ErbB-ligand binding (Carpenter and Cohen 1990). Thus it was tested if the invariant genomic organization found for the receptor-binding EGF domains for the ErbB ligands is also preserved in genomic sequences encoding a sample of unrelated EGF domains. To test this hypothesis, the proteins TGF alpha (as a reference), Epidermal Growth Factor and Notch-1 were tested. TGF alpha harbors a single EGF domain, which is responsible for receptor binding. The Epidermal Growth Factor in comparison harbors nine EGF domains; only the ninth of these being responsible for receptor binding. Notch-1 conversely is another signaling molecule that harbors thirty six EGF domains, none of these being responsible for ErbB-receptor binding (Figure 5A). The genomic sequences encoding these thee

genes were examined, in order to elucidate the genomic organization encoding their different EGF domains. For the epidermal growth factor, only the ErbB-receptor-binding EGF domain was encoded by a split codon. In contrast, the eight remaining EGF domains were wholly encoded within individual exons (Figure 5B). Conversely, for Notch-1, a heterogeneous genomic organization was observed for the thirty six encoded EGF domains (Figure 5B). Of these, only the first and the thirtieth EGF domain harbors a split exon topology at the position found for the ErbB-receptor binding domains. From these data it can be concluded that the general topology of genomic DNA encoding EGF domains in general does not necessarily require a split exon-exon structure and stop codon immediately after Exon A, as demonstrated for the ErbB-receptor binding domains in mammals.

Genes encoding ErbB ligands that do not harbor a split exon-exon structure encoding the EGF domain remain biologically active. For example, virally encoded ErbB ligands exist in nature, even though their genomes lack intronic sequences to split the EGF domain encoding region (E.g. VGF; NCBI Accession number U18337, embedded protein sequence # AAA69306). Furthermore, it is common practice in molecular biology to express genes in the form of intron-less cDNA sequences under the control of various transcriptional promoters (Maniatis et al. 1982). In this way recombinant genes encoding promoter-less ErbB ligands have been constructed, these which encode functional and active recombinant proteins (Groenen et al. 1994). Thus the evolutionary conserved exon-exon junctions found in genes encoding the different mammalian ErbB-ligands (Figure 5) are not required for the generation of functional ligands harboring the conserved six-cysteine EGF domain in mammalian cells.

The formation of functional alternative splice variants of ErbB ligands with a shortened EGF domain that ends after cysteine 4 would provide a functional explanation as to the conservation of this domain sequence. The best proof that such truncated ErbB ligand variants exist in nature is to demonstrate that such isoforms are indeed expressed. A saturation cloning effort has been performed to pull out all isoforms of the well characterized NRG1 gene. Indeed there exists a truncated NRG1 variant, which is identical to other typical NRG1 alpha isoforms, except that its sequence ends one amino acid after the fourth cysteine of the receptor-binding EGF domain (Heregulin gamma – not to be confused with gamma heregulin (Falls 2003). An examination of this protein's encoding sequence (Accession numbers NP_004486

and NM_004495) in relation to the NRG1 genomic locus, furthermore confirms that this variant sequence harbors an extended exon A, resulting in it protein's truncation (data not shown). Therefore a proof of principle that such truncated variants exist is demonstrated for NRG1.

Randomly generated transcripts provide a very poor representation of ErbB ligand sequences in public databases, such as is the case for EST sequences, particularly due to the very low expression commonly found for these genes. Nevertheless a bioinformatics search was performed to search for expressed transcripts of genes, or gene fragments, in search of truncated ErbB ligands within the EGF domain. To achieve this, the EGF domain for the different mammalian ErbB ligands (Figure 4) were used to query the NCBI NR, EST and PATENT genomic databases by method of TBLASTN, in order to search for sequences with truncated homologous sequences. These DNA sequences were extracted, and where appropriate translated into six reading frames (EMBOSS-Transeq). The relevant reading frame encoding the truncated EGF domain was chosen. Interestingly, two different classes of predicted protein sequences were discovered:

Class I: Sequences encoding a protein truncated after cysteine-4 as would be expected upon the extension of Exon A.

Class II: Sequences which encode a partial EGF domain (exon A) with alternative splice variations, in which Exon B is not encoded. The proteins encoded by this class of splice variant tends to be heterogeneous in length beyond the expression of the shortened EGF domain, depending on the alternative exon sequences that are present beyond exon A.

A list of the Class I and Class II protein sequences are shown below, inclusive of their encoded protein sequences. Unless the protein sequences were already known, the sequences provided here were translated and the appropriate reading frame encoding the truncated EGF domain was chosen. It should be noted that some of these sequences, particularly the EST sequences are partial sequences, and also are prone to occasional sequencing error. Thus, the full translated sequences are often given, regardless if an initiating methionine were noted in the translated sequence or not. These data verify the existence of two classes of ErbB ligand splice variants which encode a truncated EGF domain lacking the C-loop of the EGF domain, in a diverse range of species, including humans and other mammals, birds and fish.

Table 2 Class I variants

Sequences found in the EST, NR and Patent (DNA) databases which potentially encode ErbB ligand variants comprising an elongated Exon A, resulting in a protein sequence truncated after the conserved cysteine-4 of the EGF domain.

Sequence ID Number	Gene	Accessions	Database & Details	Species	Linked Protein Accession
140	NRG1	A81177.1	Patent WO9914323	?	?
141	NRG1	AX269478.1	Patent WO0164876	Human	?
142	NRG1	AX271009.1	Patent WO0164877	Human	?
143	NRG1	NM_004495.1	NR	Human	NP_004486
144	NRG1	AF026146.1	NR	Human	AAD01795
145	NRG1	NM_178591.1	NR	Mouse	NP_848706
146	NRG1	AK051824.1	NR (RIKEN)	Mouse	BAC34784
147	NRG1	BY212704.1	NR (RIKEN)	Mouse	None
148	NRG2	AI041451.1	EST	Human	None
149	NRG2	AX406619.1	Patent WO0222685	Human	?
150	NRG3	BX495970.1	EST	Human	None
151	NRG4	BE787057.1	EST	Human	None
152	NRG4	BF061527.1	EST	Human	None
153	NRG4	BX095400.1	EST	Human	None
154	NRG4	BB637399.1	EST	Mouse	None
155	NRG4	BB637505.1	EST	Mouse	None
156	NRG4	AI743118.1	EST	Human	None
157	NRG4	AU059620.1	EST	Pig	None
158	NRG4	C94578.1	EST	Pig	None

159	TGF alpha	AK089870.1	NR (RIKEN)	Mouse	None
160	TGF alpha	I01190.1	Patent US 4742003	Human	?
161	Epiregulin	AR019352.1	patent US 5783417	Human	?
162	Epiregulin	AR019354.1	patent US 5783417	Human	?
163	Epiregulin	AR019353.1	patent US 5783417	Mouse	?
164	Epiregulin	BC035806.1	EST (HTC)	Human	None
165	Epiregulin	BM561909.1	EST (AGENCOURT)	Human	None

Sequence ID # 85
Translation of Accession number: A81177.1
TARGAGEEFPETCWNSGLARRPGAERRRLPDDGSVSRTVITSPRSGCEGAGQR
PGREPPAAGPIDDFPGRQEQPREPGRAPVPGGRTARRVRAALPAGNGRRPRA
ARAPQRGRSLSPSRDKLFPNPIRALGPNSPAPRAVRVERSVSGEMSERKEGRG
KGKGKKKERGSGKKPESAAGSQSPALPPQLKEMKSQESAAGSKLVLRCETSS
EYSSLRFKWFKNGNELNRKNKPQNIKIQKKPGKSELRINKASLADSGEYMCK
VISKLGNDSASANITIVESNEIITGMPASTEGAYVSSESPIRISVSTEGANTSSSTS
TSTTGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCK*

Sequence ID # 86
Translation of Accession number: AX269478.1
TSTSTTGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCK*

Sequence ID # 87
Translation of Accession number: AX271009.1
TSTSTTGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCK*

Sequence ID # 88
Translation of Accession number: NM_004495.1
MSERKEGRGKGKGKKKERGSGKKPESAAGSQSPALPPQLKEMKSQESAAGS
KLVLRCETSSEYSSLRFKWFKNGNELNRKNKPQNIKIQKKPGKSELRINKASL

ADSGEYMCKVISKLGNDSASANITIVESNEIITGMPASTEGAYVSSESPIRISVS
TEGANTSSSTSTSTTGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCK*

Sequence ID # 89
Translation of Accession number: AF026146.1
MSERKEGRGKGKGKKKERGSGKKPESAAGSQSPALPPQLKEMKSQESAAGS
KLVLRCETSSEYSSLRFKWFKNGNELNRKNKPQNIKIQKKPGKSELRINKASL
ADSGEYMCKVISKLGNDSASANITIVESNEIITGMPASTEGAYVSSESPIRISVS
TEGANTSSSTSTSTTGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCK*

Sequence ID # 90
Translation of Accession number: NM_178591.1
MSERKEGRGKGKGKKKDRGSRGKPAPAEGDPSPALPPRLKEMKSQESAAGS
KLVLRCETSSEYSSLRFKWFKNGNELNRRNKPQNVKIQKKPGKSELRINKASL
ADSGEYMCKVISKLGNDSASANITIVESNDLTTGMSASTERPYVSSESPIRISVS
TEGANTSSSTSTSTTGTSHLIKCAEKEKTFCVNGGECFMVKDLSNPSRYLCK*

Sequence ID # 91
Translation of Accession number: AK051824.1
MSERKEGRGKGKGKKKDRGSRGKPAPAEGDPSPALPPRLKEMKSQESAAGS
KLVLRCETSSEYSSLRFKWFKNGNELNRRNKPQNVKIQKKPGKSELRINKASL
ADSGEYMCKVISKLGNDSASANITIVESNDLTTGMSASTERPYVSSESPIRISVS
TEGANTSSSTSTSTTGTSHLIKCAEKEKTFCVNGGECFMVKDLSNPSRYLCK*

Sequence ID # 92
Translation of Accession number: BY212704.1
MSASTERPYVSSESPIRISVSTEGANTSSSTSTSTTGTSHLIKCAEKEKTFCVNG
GECFMVKDLSNPSRYLCK*

Sequence ID # 93
Translation of Accession number: AI041451.1
TRPKLKKMKSQTGQVGEKQSLKCEAAAINPQPSYRWFKDGKELNRSRDIRIK
YGNGRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVRGRLYVNSVTTTLSS
WSGHAGKCNXTAKSYCVNGGVCYYIEGINQLSCK*

Sequence ID # 94
Translation of Accession number: AX406619.1
SSSSFDVGHEGDDSWGLGIVSVRHWHMSLIPSVSTTLSSWSGHARKCNETAK
SYCVNGGVCYYIEGINQLSCK*

Sequence ID # 95

Translation of Accession number: BX495970.1
EINIIIWYYFPSAWRTCFNISSSVGLLLTNSYKFYTTTYSTERSEHFKPCRDKDL AYCLNDGECFVIETLTGSHKHCR*

Sequence ID # 96
Translation of Accession number: BE787057.1
NYLQIKMPTDHEEPCGPSHKSFCLNGGLCYVIPTIPSPFCRK

Sequence ID # 97
Translation of Accession number: BF061527.1
MPTDHEEPCGPSHKSFCLNGGLCYVIPTIPSPFCRK*

Sequence ID # 98
Translation of Accession number: BX095400.1
MPTDHEEPCGPSHKSFCLNGGLCYVIPTIPSPFCRK*

Sequence ID # 99
Translation of Accession number: BB637399.1
MPTGNFLSRAALWSQAQVILPQWGDLLCDPYYPQPIL*

Sequence ID # 100
Translation of Accession number: BB637505.1
MPTGNFLSRAALWSQAQVILPQWGDLLCDPYYPQPIL*

Sequence ID # 101
Translation of Accession number: AI743118.1
SHKSFCLNGGLCYVIPTIPSPFCRK*

Sequence ID # 102
Translation of Accession number: AU059620.1
EPCGPSHRSFCLNGGICYVIPTIPSPFCRK*

Sequence ID # 103
Translation of Accession number: C94578.1
EPCGPSHRSFCLNGGICYVIPTIPSPFCRK*

Sequence ID # 104
Translation of Accession number: AK089870.1
CLFAPADSPVAAAVVSHFNKCPDSHTQYCFHGTCRFLVQEEKPACV

Sequence ID # 105
Translation of Accession number: I01190.1
DLSPASFLSPADPPVAAAVVSHFNDCPDSHTQFCFHGTCRFLVQEDKPACV*

Sequence ID # 106
Translation of Accession number: AR019352.1
VQTEDNPRVAQVSITKCSSDMNGYCLHGQCIYLVDMSQNYCR

Sequence ID # 107
Translation of Accession number: AR019354.1
QTEDNPRVAQVSITKCSSDMNGYCLHGQCIYLVDMSQNYC

Sequence ID # 108
Translation of Accession number: AR019353.1
VQMEDDPRVAQVQITKCSSDMDGYCLHGQCIYLVDMREKFCR

Sequence ID # 109
Translation of Accession number: BC035806.1
MTAGRRMEMLCAGRVPALLLCLGFHLLQAVLSTTVIPSCIPGESSDNCTALVQ
TEDNPRVAQVSITKCSSDMNGYCLHGQCIYLVDMSQNYCR*

Sequence ID # 110
Translation of Accession number: BM561909.1
MTAGRRMEMLCAGRVPALLLCLGFHLLQAVLSTTVIPSCIPGESSDNCTALVQ
TEDNPRVAQVSITKCSSDMNGYCLHGQCIYLVDMSQNYCR*

Table 3: Class II variants

Sequences found in the EST, NR and Patent (DNA) databases potentially encode ErbB ligands which include Exon A but lack Exon B, resulting in the predicted expression of proteins of varying lengths extending beyond that of a shortened EGF domain (to the conserved Cys-4).

Sequence number	Gene	Accession	Database & Details	Species	Linked Protein Accession
166	NRG2	AA706226.1	EST	Human	None
167	NRG2	BX089049.1	EST	Human	None
168	NRG2	AI152190.1	EST	Mouse	None
169	NRG2	AL918370.1	EST	Zebrafish	None

170	NRG3	BU465274.1	EST	Chicken	None
171	NRG4	BU372401.1	EST	Chicken	None
172	NRG4	BE624667.1	EST	Mouse	None
173	Amphiregulin	BE064716.1	EST	Human	None
174	Betacellulin	BG194271.1	EST (RAGE)	Human	None
175	BY735030.1	BY735030.1	EST (RIKEN)	Mouse	None
176	HB-EGF	X89728.1	NR	Cercopithecus aethiops (African green monkey)	CAA61880
177	Epigen	BD274363.1	Patent JP 2002530064-A/7.	Human	?
178	Epigen	AX261946.1	Patent WO0172781	Human	?
179	Epigen	AX261991.1	Patent WO0172781	Human	?
180	Epigen	BD274361.1	Patent JP 2002530064-A/5.	Human	?
181	Epigen	BD209747.1	Patent JP 2002512798-A/219	Human	?
182	Epigen	BD274362.1	Patent JP 2002530064-A/6.	Human	?

<u>Translated sequences</u>:

Sequence ID # 111
Translation of Accession number: AA706226.1

PGEKATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRS
RDIRIKYGNGRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGGRLYVNSVT
TTLSSWSGHARKCNXTAKSYCVNGGVCYYIEGINQLSCKAPGLHCLELGTQS
HHFPISASPGSSQGSWNQLPQHPLS

Sequence ID # 112
Translation of Accession number: BX089049.1

EAENILGKDTVRXRLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYYIE GINQLSCKAHGLHCLELGTQSHHFPISASPGSSQGSWNQLPQHPLSALGGEGS PGGDAVRTPGPQSCA

Sequence ID # 113
Translation of Accession number: AI152190.1

VRQRRETPSPPIAGSRMARNSTGVVIFASSMAMAVSTTLSSWSGHARKCNET AKSYCVNGGVCYYIEGINQLSCKG*

Sequence ID # 114
Translation of Accession number: AL918370.1

KDCASAPKVKPMDSQWLQEGKKLTLKCEAVGNPSPSFNWYKDGSQLRQKK TVKIKTNKKNSKLHISKVRLEDSGNYTCVVENSLGRENATSFVSVQSITTTLSP GSSHARKCNETEKTYCINGGDCYFIHGINQLSCKCPNDYTGERCQTSVMAGF YKAEELYQNEC*

Sequence ID # 115
Translation of Accession number: BU465274.1

AVQSLELLQQTWRLSTLQFEYDRRVACGFHYTTTYSTERSEHFKPCKDKDLA YCLNEGECFVIETLTGSHKHCRSNCPSGVFCW*

Sequence ID # 116
Translation of Accession number: BU372401.1

MRTDHEELCGTSYGSFCLNGGICYMIPTVPSPFCRHLPKAANQASALHKSVFS IFVLHTDTTALPSCHLMPAHFYTQ*

Sequence ID # 117
Translation of Accession number: BE624667.1

MPTDHEQPCGPRHRSFCLNGGICIDPYYPHPFCRFYHLFLRHCLLKPFVQLGTL VYPVFLKELFH*

Sequence ID # 118
Translation of Accession number: BE064716.1

DVIAQHKPESENTSDKPKRKKKGGKNGKNRRNRKKKNPCDAEFQNFCIHGE CKYIEHLEAVTCNVSRIFP*

Sequence ID # 119
Translation of Accession number: BG194271.1

LXATTQSKWKGHSSRCPKQYKHYCIKGRCRFVVAEQTPSCVPLRKRRKRKK KEEEMETLGKDMTPINEDIEETNIAYKAMKLPPGWWQAAKCLAHLKMDRM RLRKTASRHEF*

Sequence ID # 120
Translation of Accession number: BY735030.1

KSLTWKSFNFLSLLLPLGSTGTRRILCPLSTPSCSAGLAILHCVVADGNTTRTP
ETNGSLCGAPGENCTGTTPRQKVKTHFSRCPKQYKHYCIHGRCRFVVDEQTP
SCMARLSIYLWRN*

Sequence ID # 121
Translation of Accession number: X89728.1
MKLLPSVVLKLLLAAVLSALVTGESLEQLRRGPAAGTSNPDPSTGSTDQLLRL
GGGRDRKVRDLQEADLDLLRVTLSSKPQALATPSKEEHGKRKKKGKGLGKK
RDPCLRKYKDFCIHGECKYVKELRAPSCMAAGQKDVT

Sequence ID # 122
Translation of Accession number: BD274363.1
MTALTEEAAVTVTPPITAQQADNIEGPIALKFSHLCLEDHNSYCINGACAFHH
ELEKAICRCLKLKSPYNVCSGERRPL*

Sequence ID # 123
Translation of Accession number: AX261946.1
GTREALCYRCFCPLNTAMRALTEEAAVTVTPPITAQQADNIEGPIALKFSHLC
LEDHNSYCINGACAFHHELEKAICRCLKLKSPYNVCSGERRPL*

Sequence ID # 124
Translation of Accession number: AX261991.1
GTREALCYRCFCPLNTAMRALTEEAAVTVTPPITAQQADNIEGPIALKFSHLC
LEDHNSYCINGACAFHHELEKAICRCLKLKSPYNVCSGERRPL*

Sequence ID # 125
Translation of Accession number: BD274361.1
LQEMALGVPISVYLLFNAMTALTEEAAVTVTPPITAQQADNIEGPIALKFSHLC
LEDHNSYCINGACAFHHELEKAICRCLKLKSPYNVCSGERRPL*

Sequence ID # 126
Translation of Accession number: BD209747.1
KDKRKKVKQLQEMALGVPISVYLLFNAMTALTEEAAVTVTPPITAQQGNWT
VNKTEADNIEGPIALKFSHLCLEDHNSYCINGACAFHHELEKAICRCLKLKSPY
NVCSGERRPL*

Sequence ID # 127
Translation of Accession number: BD274362.1
MALGVPISVYLLFNAMTALTEEAAVTVTPPITAQQADNIEGPIALKFSHLCLED
HNSYCINGACAFHHELEKAICRCLKLKSPYNVCSGERRPL

DNA sequences encoding truncated class 1 variants (Figure 4):

Sequence ID # 128
ACTGGGACAAGCCATCTTGTAAAATGTGCGGAGAAGGAGAAAACTTTCTGTGTGAATGGAGGGGAGTGC
TTCATGGTGAAAGACCTTTCAAACCCCTCGAGATACTTGTGCAAGTAA

Sequence ID # 129
TCCTGGTCGGGGCACGCCCGGAAGTGCAACGAGACAGCCAAGTCCTATTGCGTCAATGGAGGCGTCTGC
TACTACATCGAGGGCATCAACCAGCTCTCCTGCAAGTAA

Sequence ID # 130
GAGCGATCCGAGCACTTCAAACCCTGCCGAGACAAGGACCTTGCATACTGTCTCAATGATGGCGAGTGC
TTTGTGATCGAAACCCTGACCGGATCCCATAAACACTGTCGGTAA

Sequence ID # 131
GATCACGAAGAGCCCTGTGGTCCCAGTCACAAGTCGTTTTGCCTGAATGGGGGGCTTTGTTATGTGATA
CCTACTATTCCCAGCCCATTTTGTAGGTGA

Sequence ID # 132
TCCGTAAGAAATAGTGACTCTGAATGTCCCCTGTCCCACGATGGGTACTGCCTCCATGATGGTGTGTGC
ATGTATATTGAAGCATTGGACAAGTATGCATGCAAGTAA

Sequence ID # 133
GCAGTGGTGTCCCATTTTAATGACTGCCCAGATTCCCACACTCAGTTCTGCTTCCATGGAACCTGCAGG
TTTTTGGTGCAGGAGGACAAGCCAGCATGTGTGTAA

Sequence ID # 134
AAGCGGAAAGGCCACTTCTCTAGGTGCCCCAAGCAATACAAGCATTACTGCATCAAAGGGAGATGCCGC
TTCGTGGTGGCCGAGCAGACGCCCTCCTGTGTGTAA

Sequence ID # 135
AGAAACAGAAAGAAGAAAAATCCATGTAATGCAGAATTTCAAAATTTCTGCATTCACGGAGAATGCAAA
TATATAGAGCACCTGGAAGCAGTAACATGCAAGTAA

Sequence ID # 136
GGGCTAGGGAAGAAGAGGGACCCATGTCTTCGGAAATACAAGGACTTCTGCATCCATGGAGAATGCAAA
TATGTGAAGGAGCTCCGGGCTCCCTCCTGCATGTAA

Sequence ID # 137
GTGGCTCAAGTGTCAATAACAAAGTGTAGCTCTGACATGAATGGCTATTGTTTGCATGGACAGTGCATC
TATCTGGTGGACATGAGTCAAAACTACTGCAGGTAA

Sequence ID # 138
GTAGCTCTGAAGTTCTCTCATCCTTGTCTGGAAGACCATAATAGTTACTGCATTAATGGAGCATGTGCA
TTCCACCATGAGCTGAAGCAAGCCATTTGCAGGTAA

Sequence ID # 139
ATAGCCTTGAAGTTCTCACACCTTTGCCTGGAAGATCATAACAGTTACTGCATCAACGGTGCTTGTGCA
TTCCACCATGAGCTAGAGAAAGCCATCTGCAGGTAA

Novel splice variants of ErbB ligands

Currently preferred embodiments according to the present invention include isolated polynucleotides selected from the following:

1. Polynucleotides encoding the extended EGF domain derived directly from genomic data (denoted herein as Class 1): namely SEQ ID NOS:128 to 139.

2. Polynucleotides encoding Class 1 variants or fragments of variants derived from the EST, NR and Patent databases (Table 2 excluding gamma variants): namely SEQ ID NOS:148 to 165.

3. Polynucleotides encoding Class 2 variants of fragments of variants derived from the EST, NR and Patent databases (Table 3): namely SEQ ID NOS:166 to 182.

It is explicitly understood that all known sequences are excluded from the scope of the present invention.

Currently preferred embodiments according to the present invention include polypeptides comprising the following:

1. Polypeptides comprising truncated EGF domain derived directly from genomic data (denoted herein Class 1) namely SEQ ID NOS:73 to 84.

2. Class 1 variants or fragments of variants derived from the EST, NR and Patent databases (translation of Table 2 sequences excluding gamma variants) namely SEQ ID NOS:93 to 110.

3. Class 2 variants of fragments of variants derived from the EST, NR and Patent databases (translated sequences of Table 3) namely SEQ ID NOS:111 to 127.

It is explicitly understood that all known sequences are excluded from the scope of the present invention.

Thus, according to one aspect of the present invention there are provided isolated nucleic acids comprising a genomic, complementary or composite polynucleotide sequence encoding a polypeptide being capable of binding to a mammalian ErbB which is at least 70%, preferably at least 80%, more preferably at least 90% or more, say at least 95%, or 100% homologous (similar+identical acids) to SEQ ID NOS:73-84 and SEQ ID NOS:93-127. Homology is determined for example using Gapped BLAST-based searches (Altschul et. al. 1997) with preferred matrix BLOSUM62 (protein-based searches) and the following default parameters as defined by the NCBI BLAST web site:

-G Cost to open gap [Integer]
default = 5 for nucleotides 11 proteins
-E Cost to extend gap [Integer]
default = 2 nucleotides 1 proteins
-q Penalty for nucleotide mismatch [Integer]
default = -3
-r reward for nucleotide match [Integer]
default = 1
-e expect value [Real]
default = 10
-W wordsize [Integer]
default = 11 nucleotides 3 proteins
-y Dropoff (X) for blast extensions in bits (default if zero)
default = 20 for blastn 7 for other programs
-X X dropoff value for gapped alignment (in bits)
default = 15 for al programs except for blastn for which it does not apply
-Z final X dropoff value for gapped alignment (in bits)
50 for blastn 25 for other programs

Accordingly, any nucleic acid sequence which encodes the amino acid sequence of an ErbB ligand can be used to produce recombinant molecules which express this ligand. In particular embodiments, the polynucleotide according to another aspect of the present invention encodes a polypeptide as set forth in SEQ ID NOS:73 to 84 and SEQ ID NOS:93 to 127, or a portion thereof, which retains at least one biological, immunological or other functional characteristic or activity of a known ligand of at least one ErbB receptor.

The EGF-encoded variant domains disclosed herein comprise a consensus sequence that may be represented as follows: (X-8)-C-(X-7)-C-(X-2 to 3)-G-X-C-(X-10 to 13)–C-X, wherein X is any amino acid. This is the consensus pattern presented in Figure 4. Shorter or longer amino-terminal sequences (X-8 hereinabove) can provide or define biological activity. Generally, synthetic peptides derived from the novel ligands may have extensions including an amino-terminal tail of amino acids.

It is to be understood that the present invention encompasses all fragments or variants including such amino terminal extensions, with the proviso that the C loop of the EGF domain is absent from these derivatives.

Methods for DNA sequencing are well known and generally available in the art, and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (U.S. Biochemical Corp, Cleveland, Ohio), Taq polymerase (Perkin

Elmer), thermostable T7 polymerase (Amersham, Chicago, Ill.), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, Md.). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, Nev.), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, Mass.) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding ErbB ligand isoforms, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring ErbB ligand isoforms, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode ErbB ligand isoforms and their variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring ErbB ligand isoforms under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding ErbB ligand isoforms or their derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding ErbB ligand isoforms and their derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences, or fragments thereof, which encode ErbB ligand isoforms and their derivatives, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding ErbB ligand isoforms or any fragment thereof.

The present invention also includes polynucleotide sequences that are capable of hybridizing to the nucleotide sequences according to the present invention. According to one embodiment, the polynucleotide is preferably hybridizable with SEQ ID NOS: 73 to 84 and 93 to 127.

Hybridization for long nucleic acids (e.g., about 200 bp in length) is effected according to preferred embodiments of the present invention by stringent or moderate hybridization, wherein stringent hybridization is effected by a hybridization solution containing 10% dextran sulfate, 1 M NaCl, 1% SDS and $5x10^6$ rpm ^{32}P labeled probe, at 65°C, with a final wash solution of 0.2xSSC and 0.1% SDS and final wash at 65°C; whereas moderate hybridization is effected by a hybridization solution containing 10% dextrane sulfate, 1 M NaCl, 1% SDS and $5x10^6$ cpm ^{32}P labeled probe, at 65°C, with a final wash solution of 1xSSC and 0.1% SDS and final wash at 50°C.

According to preferred embodiments the polynucleotide according to this aspect of the present invention is as set forth in SEQ ID Nos:73 to 84 and 93 to 127, or a portion thereof, said portion preferably encodes a polypeptide comprising an amino acid stretch of at least 80%, preferably at least 85%, more preferably at least 90% or more, most preferably 95% or more identical to positions the polynucleotide sequence encoding the truncated ErbB receptor-binding EGF domain devoid of the C-loop.

According to still another embodiment of the present invention there is provided an oligonucleotide of at least 17, at least 18, at least 19, at least 20, at least 22, at least 25, at least 30 or at least 40, bases specifically hybridizable with the isolated nucleic acid described herein.

Hybridization of shorter nucleic acids (below 200 bp in length, e.g., 17-40 bp in length) is effected by stringent, moderate or mild hybridization, wherein stringent hybridization is effected by a hybridization solution of 6xSSC and 1% SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5% SDS, 100 µg/ml denatured salmon sperm DNA and 4.1% nonfat dried milk, hybridization temperature of 1-1.5°C below the T_m, final wash solution of 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 m EDTA (pH 7.6), 0.5% SDS at 1-1.5°C below the T_m. Moderate hybridization is effected by a hybridization solution of 6xSSC and 0.1% SDS or 3 M TMACI, 0.01 M sodium phosphate(pH 6.8), 1 mM EDTA (pH 7.6), 0.6% SDS, 100 µg/ml denatured salmon sperm DNA and 0.1% nonfat dried milk, hybridization temperature of 2-2.5°C below the T_m, final wash solution of 3 M

TMACl, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5% SDS at 1-1.5°C below the T_m, final wash solution of 6xSSC, and final wash at 22°C; whereas mild hybridization is effected by a hybridization solution of 6xSSC and 1% SDS or 3M TMACl, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5% SDS, 100 μg/ml denatured salmon sperm DNA and 0.1% nonfat dried milk, hybridization temperature of 37°C, final wash solution of 6xSSC and final wash at 22°C.

According to an additional aspect of the present invention there is provided a pair of oligonucleotides each independently of at least 17-40 bases specifically hybridizable with the isolated nucleic acid described herein in an opposite orientation so as to direct exponential amplification of a portion thereof, say of 50 to 2000 bp, in a nucleic acid amplification reaction, such as a polymerase chain reaction. The polymerase chain reaction and other nucleic acid amplification reactions are well known in the art and require no further description herein. The pair of oligonucleotides according to this aspect of the present invention are preferably selected to have comparable melting temperatures (T_m), e.g., melting temperatures which differ by less than that 7°C, preferably less than 5°C, more preferably less than 4°C, most preferably less than 3°C, ideally between 3°C and 0°C. Consequently, according to yet an additional aspect of the present invention there is provided a nucleic acid amplification product obtained using the pair of primers described herein. Such a nucleic acid amplification product can be isolated by gel electrophoresis or by any other size-based separation technique. Alternatively, such a nucleic acid amplification product can be isolated by affinity separation, either stranded affinity or sequence affinity. In addition, once isolated, such a product can be further genetically manipulated by restriction, ligation and the like, to serve any one of a plurality of applications associated with regulation of ErbB activity as further detailed herein.

The nucleic acid sequences encoding ErbB ligand isoforms may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). In particular, genomic DNA is first amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same

linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). The primers may be designed using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, Minn.), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°C to72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may also be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR.

Another method which may be used to retrieve unknown sequences is that of Parker, J. D. et al. (1991; Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, Calif.). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide)

which are laser activated, and detection of the emitted wavelengths by a charge coupled devise camera. Output/light intensity may be converted to electrical signal using appropriate software (e.g. Genotyper™ and Sequence Navigator™, Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

Thus, this aspect of the present invention encompasses (i) polynucleotides as set forth in SEQ ID NOs: DNA sequence IDs claimed (exclusive of the known gamma isoform):128 to 139 and 148 to 192; (ii) fragments thereof; (iii) sequences hybridizable therewith; (iv) sequences homologous thereto; (v) sequences encoding similar polypeptides with different codon usage; (vi) altered sequences characterized by mutations, such as deletion, insertion or substitution of one or more nucleotides, either naturally occurring or man induced, either randomly or in a targeted fashion.

Producing the novel variants

Synthetic peptides comprising the novel variants

Peptides were synthesized on an Applied Biosystems (ABI) 430A peptide synthesizer using standard *tert*-butyloxycarbonyl (*t*-Boc) chemistry protocols as provided (version 1.40; *N*-methylpyrrolidone/hydroxybenzotriazole). Acetic anhydride capping was employed after each activated ester coupling. The peptides were assembled on phenylacetamidomethyl polystyrene resin using standard side chain protection except for the use of *t*-Boc Glu(*O*-cyclohexyl) and *t*-Boc Asp(*O*-cyclohexyl). The peptides were deprotected using the ``Low-High" hydrofluoric acid (HF) method of Tam *et al.* (23) In each case crude HF product was purified by reverse phase HPLC (C-18 Vydac, 22x250 mm), diluted without drying into folding buffer (1 M urea, 100 mM Tris, pH 8.0, 1.5 mM oxidized glutathione, 0.75 mM reduced glutathione, 10 mM Met), and stirred for 48 h at 4 °C. Folded, fully oxidized peptides were purified from the folding mixture by reverse phase HPLC and characterized by electrospray mass spectroscopy; quantities were determined by amino acid analysis.

Constructs comprising the novel variants

According to another aspect of the present invention there is provided a nucleic acid construct comprising the isolated nucleic acid described herein.

According to a preferred embodiment the nucleic acid construct according to this aspect of the present invention further comprising a promoter for regulating the expression of the isolated nucleic acid in a sense or antisense orientation. Such promoters are known to be cis-acting sequence elements required for transcription as they serve to bind DNA dependent RNA polymerase which transcribes sequences present downstream thereof. Such down stream sequences can be in either one of two possible orientations to result in the transcription of sense RNA which is translatable by the ribosome machinery or antisense RNA which typically does not contain translatable sequences, yet can duplex or triplex with endogenous sequences, either mRNA or chromosomal DNA and hamper gene expression, all as is further detailed hereinunder.

While the isolated nucleic acid described herein is an essential element of the invention, it is modular and can be used in different contexts. The promoter of choice that is used in conjunction with this invention is of secondary importance, and will comprise any suitable promoter sequence. It will be appreciated by one skilled in the art, however, that it is necessary to make sure that the transcription start site(s) will be located upstream of an open reading frame. In a preferred embodiment of the present invention, the promoter that is selected comprises an element that is active in the particular host cells of interest. These elements may be selected from transcriptional regulators that activate the transcription of genes essential for the survival of these cells in conditions of stress or starvation, including the heat shock proteins.

Vectors and host cells

In order to express a biologically active ErbB ligand isoform, the nucleotide sequences encoding ErbB ligand isoforms or functional equivalents according to the present invention may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Vectors can be introduced into cells or tissues by any one of a variety of known methods within the art, including in vitro recombinant DNA techniques, synthetic

techniques, and in vivo genetic recombination. Such methods are generally described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Springs Harbor Laboratory, New York 1989, 1992;, in Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, Baltimore, Md. 1989;, Chang et al., Somatic Gene Therapy, CRC Press, Ann Arbor, Mich. 1995; Vega et al., Gene Targeting, CRC Press, Ann Arbor Mich. 1995; Vectors: A Survey of Molecular Cloning Vectors and Their Uses, Butterworths, Boston Mass. 1988; and Gilboa et al. (1986) Biotechniques 4 (6): 504-512, and include, for example, stable or transient transfection, lipofection, electroporation and infection with recombinant viral vectors. In addition, see U.S. Patent No. 4,866,042 for vectors involving the central nervous system and also U.S. Patent Nos. 5,464,764 and 5,487,992 for positive-negative selection methods.

A variety of expression vector/host systems may be utilized to contain and express sequences encoding ErbB ligand isoforms. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed. The expression of the construct according to the present invention within the host cell may be transient or it may be stably integrated in the genome thereof.

The polynucleotides of the present invention may be employed for producing polypeptides by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

The "control elements" or "regulatory sequences" are those non-translated regions of the vector- enhancers, promoters, 5' and 3' untranslated regions - which interact with host cellular proteins to carry out transcription and translation. Such

elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, LaJolla, Calif.) or pSport1™ plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding variant ErbB-ligand, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for variant ErbB-ligand expression. For example, when large quantities of variant ErbB-ligand are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding variant ErbB-ligand may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S. M. Schuster (1989) J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) Methods Enzymol. 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding variant ErbB-ligand may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196.

An insect system may also be used to express variant ErbB-ligand. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding variant ErbB-ligand may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of variant ErbB-ligand will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which variant ErbB-ligand may be expressed (Engelhard, E. K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding variant ErbB-ligand may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing variant ErbB-ligand in infected host cells (Logan, J. and Shenk, T. (1984) Proc. Natl. Acad. Sci. 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of 6 to

10M are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding variant ErbB-ligand. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding variant ErbB-ligand, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Polypeptide purification

Host cells transformed with nucleotide sequences encoding ErbB ligand isoforms may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. The polynucleotide encoding for ErbB ligand isoforms may include a signal peptide which direct secretion of ErbB ligand isoforms through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding ErbB ligand isoforms to nucleotide sequences encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAG extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences, such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, Calif.), between the purification domain and the ErbB ligand isoforms encoding sequence may be used to facilitate

purification. One such expression vector provides for expression of a fusion protein containing ErbB ligand isoforms and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on immobilized metal ion affinity chromatography. (IMIAC) (See, e.g., Porath, J. et al. (1992) Prot. Exp. Purif. 3:263-281.) The enterokinase cleavage site provides a means for purifying ErbB ligand isoforms from the fusion protein. (See, e.g., Kroll, D. J. et al. (1993) DNA Cell Biol. 12:441-453.)

Fragments of ErbB ligand isoforms may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, T. E. (1984) Protein: Structures and Molecular Properties, pp. 55-60, W. H. Freeman and Co. New York, N.Y.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431A peptide synthesizer (Perkin Elmer). Various fragments of ErbB ligand isoforms may be synthesized separately and then combined to produce the full length molecule.

Transgenic animals or cell lines

The present invention has the potential to provide transgenic gene and polymorphic gene animal and cellular (cell lines) models as well as for knock-out and knock-in models. These models may be constructed using standard methods known in the art and as set forth in U.S. Patent Nos. 5,487.992, 5,464,764, 5,387,742, 5,360,735, 5,347,075, 5,298,422, 5,288,846, 5,221,778, 5,175,385, 5,175,384, 5,175,383, 4,736,866 as well as Burke and Olson (1991) Methods in Enzymology, 194:251-270; Capecchi (1989) Science 244:1288-1292; Davies et al. (1992) Nucleic Acids Research, 20 (11) 2693-2698; Dickinson et al. (1993) Human Molecular Genetics, 2(8): 1299-1302; Duff and Lincoln, "Insertion of a pathogenic mutation into a yeast artificial chromosome containing the human APP gene and expression in ES cells", Research Advances in Alzheimer's Disease and Related Disorders, 1995; Huxley et al. (1991) Genomics, 9:7414 750 1991; Jakobovits et al. (1993) Nature, 362:255-261; Lamb et al.(1993) Nature Genetics, 5: 22-29; Pearson and Choi, (1993) Proc. Natl. Acad. Sci. USA 90:10578-82; Rothstein, (1991) Methods in Enzymology, 194:281-301; Schedl et al. (1993) Nature, 362: 258-261; Strauss et al. (1993) Science,

259;1904-1907. Further, patent applications WO 94/23049, WO 93/14200, WO 94/06408, WO 94/28123 also provide information.

All such transgenic gene and polymorphic gene animal and cellular (cell lines) models and knockout or knock-in models derived from claimed embodiments of the present invention, constitute preferred embodiments of the present invention.

Gene therapy

Gene therapy as used herein refers to the transfer of genetic material (e.g., DNA or RNA) of interest into a host to treat or prevent a genetic or acquired disease or condition or phenotype. The genetic material of interest encodes a product (e.g., a protein, polypeptide, peptide, functional RNA, antisense) whose production in vivo is desired. For example, the genetic material of interest can encode a ligand, hormone, receptor, enzyme, polypeptide or peptide of therapeutic value. For review see, in general, the text "Gene Therapy" (Advanced in Pharmacology 40, Academic Press, 1997).

Two basic approaches to gene therapy have evolved: (i) ex vivo and (ii) in vivo gene therapy. In ex vivo gene therapy cells are removed from a patient, and while being cultured are treated in vitro. Generally, a functional replacement gene is introduced into the cell via an appropriate gene delivery vehicle/method (transfection, transduction, homologous recombination, etc.) and an expression system as needed and then the modified cells are expanded in culture and returned to the host/patient. These genetically reimplanted cells have been shown to express the transfected genetic material in situ.

In in vivo gene therapy, target cells are not removed from the subject. Rather, the genetic material to be transferred is introduced into the cells of the recipient organism in situ, that is within the recipient. In an alternative embodiment, if the host gene is defective, the gene is repaired in situ (Culver, 1998. (Abstract) Antisense DNA & RNA based therapeutics, February 1998, Coronada, Calif.). These genetically altered cells have been shown to express the transfected genetic material in situ. The gene expression vehicle is capable of delivery/transfer of heterologous nucleic acid into a host cell. The expression vehicle may include elements to control targeting, expression and transcription of the nucleic acid in a cell selective manner as is known

in the art. It should be noted that often the 5'UTR and/or 3'UTR of the gene may be replaced by the 5'UTR and/or 3'UTR of the expression vehicle. Therefore, as used herein the expression vehicle may, as needed, not include the 5'UTR and/or 3'UTR of the actual gene to be transferred and only include the specific amino acid coding region.

The expression vehicle can include a promoter for controlling transcription of the heterologous material and can be either a constitutive or inducible promoter to allow selective transcription. Enhancers that may be required to obtain necessary transcription levels can optionally be included. Enhancers are generally any nontranslated DNA sequences which work contiguously with the coding sequence (in cis) to change the basal transcription level dictated by the promoter. The expression vehicle can also include a selection gene as described hereinbelow.

Vectors useful in gene therapy

As described herein above, vectors can be introduced into host cells or tissues by any one of a variety of known methods within the art.

Introduction of nucleic acids by infection offers several advantages over the other listed methods. Higher efficiency can be obtained due to their infectious nature. Moreover, viruses are very specialized and typically infect and propagate in specific cell types. Thus, their natural specificity can be used to target the vectors to specific cell types in vivo or within a tissue or mixed culture of cells. Viral vectors can also be modified with specific receptors or ligands to alter target specificity through receptor mediated events.

A specific example of DNA viral vector introducing and expressing recombination sequences is the adenovirus-derived vector Adenop53TK. This vector expresses a herpes virus thymidine kinase (TK) gene for either positive or negative selection and an expression cassette for desired recombinant sequences. This vector can be used to infect cells that have an adenovirus receptor which includes most cancers of epithelial origin as well as others. This vector as well as others that exhibit similar desired functions can be used to treat a mixed population of cells and can include, for example, an in vitro or ex vivo culture of cells, a tissue or a human subject.

Features that limit expression to particular cell type can also be included. Such

features include, for example, promoter and regulatory elements that are specific for the desired cell type.

In addition, recombinant viral vectors are useful for in vivo expression of a desired nucleic acid because they offer advantages such as lateral infection and targeting specificity. Lateral infection is inherent in the life cycle of, for example, retrovirus and is the process by which a single infected cell produces many progeny virions that bud off and infect neighboring cells. The result is that a large area becomes rapidly infected, most of which was not initially infected by the original viral particles. This is in contrast to vertical-type of infection in which the infectious agent spreads only through daughter progeny. Viral vectors can also be produced that are unable to spread laterally. This characteristic can be useful if the desired purpose is to introduce a specified gene into only a localized number of targeted cells.

As described above, viruses are very specialized infectious agents that have evolved, in many cases, to elude host defense mechanisms. Typically, viruses infect and propagate in specific cell types. The natural specificity of viral vectors is utilized to specifically target predetermined cell types and thereby introduce a recombinant gene into the infected cell. The vector to be used in the methods of the invention will depend on desired cell type to be targeted and will be known to those skilled in the art. For example, if breast cancer is to be treated then a vector specific for such epithelial cells would be used. Likewise, if diseases or pathological conditions of the hematopoietic system are to be treated, then a viral vector specific for blood cells and their precursors, preferably for the specific type of hematopoietic cell, would be used.

Retroviral vectors can be constructed to function either as infectious particles or to undergo only a single initial round of infection. In the former case, the genome of the virus is modified so that it maintains all the necessary genes, regulatory sequences and packaging signals to synthesize new viral proteins and RNA. Once these molecules are synthesized, the host cell packages the RNA into new viral particles, which are capable of undergoing further rounds of infection. The vector's genome is also engineered to encode and express the desired recombinant gene. In the case of non-infectious viral vectors, the vector genome is usually mutated to destroy the viral packaging signal that is required to encapsulate the RNA into viral particles. Without such a signal, any particles that are formed will not contain a genome and therefore cannot proceed through subsequent rounds of infection. The specific type of vector

will depend upon the intended application. The actual vectors are also known and readily available within the art or can be constructed by one skilled in the art using well-known methodology.

The recombinant vector can be administered in several ways. If viral vectors are used, for example, the procedure can take advantage of their target specificity and consequently, they do not have to be administered locally at the diseased site. However, when local administration can provide a quicker and more effective treatment, administration can also be performed by, for example, intravenous or subcutaneous injection into the subject. Injection of the viral vectors into a spinal fluid can also be used as a mode of administration. Following injection, the viral vectors will circulate until they recognize cells with appropriate target specificity for infection.

Thus, according to an alternative embodiment, the nucleic acid construct according to the present invention further includes a positive and a negative selection markers and may therefore be employed for selecting for homologous recombination events, including, but not limited to, homologous recombination employed in knock-in and knockout procedures. One ordinarily skilled in the art can readily design a knockout or knock-in constructs including both positive and negative selection genes for efficiently selecting transfected embryonic stem cells that underwent a homologous recombination event with the construct.

Such cells can be introduced into developing embryos to generate chimeras, the offspring thereof can be tested for carrying the knockout or knock-in constructs. Knockout and/or knock-in constructs according to the present invention can be used to further investigate the functionality of ErbB ligand isoforms. Such, constructs can also be used in somatic and/or germ cells gene therapy to increase/decrease the activity of ErbB signaling, thus regulating ErbB related responses. Further detail relating to the construction and use of knockout and knock-in constructs can be found in Fukushige, S. and Ikeda, J. E. (1996) DNA Res 3:73-50; Bedell, M. A. et al. (1997) Genes and Development 11:1-11; Bermingham, J. J. et al. (1996) Genes Dev 10:1751-1762, which are incorporated herein by reference as if set forth herein.

__Antisense__

According to still an additional aspect of the present invention there is provided an

antisense oligonucleotide comprising a polynucleotide or a polynucleotide analog of at least 10 bases, preferably between 10 and 15, more preferably between 5 and 20 bases, most preferably, at least 17- 40 bases being hybridizable in vivo, under physiological conditions, with a portion of a polynucleotide strand encoding a polypeptide at least 80%, preferably at least 85%, more preferably at least 90% or more, most preferably at least 95% or more homologous (similar+identical acids) to the sequence of the ErbB receptor- binding EGF ligand devoid of the C-loop disclosed by the present invention. Such antisense oligonucleotides can be used to downregulate expression as further detailed hereinunder. Such an antisense oligonucleotide is readily synthesizable using solid phase oligonucleotide synthesis.

The ability of chemically synthesizing oligonucleotides and analogs thereof having a selected predetermined sequence offers means for down-modulating gene expression. Three types of gene expression modulation strategies may be considered.

At the transcription level, antisense or sense oligonucleotides or analogs that bind to the genomic DNA by strand displacement or the formation of a triple helix, may prevent transcription. At the transcript level, antisense oligonucleotides or analogs that bind target mRNA molecules lead to the enzymatic cleavage of the hybrid by intracellular RNase H. In this case, by hybridizing to the targeted mRNA, the oligonucleotides or oligonucleotide analogs provide a duplex hybrid recognized and destroyed by the RNase H enzyme. Alternatively, such hybrid formation may lead to interference with correct splicing. As a result, in both cases, the number of the target mRNA intact transcripts ready for translation is reduced or eliminated. At the translation level, antisense oligonucleotides or analogs that bind target mRNA molecules prevent, by steric hindrance binding of essential translation factors (ribosomes), to the target mRNA a phenomenon known in the art as hybridization arrest, disabling the translation of such mRNAs.

Thus, antisense sequences, which as described hereinabove may arrest the expression of any endogenous and/or exogenous gene depending on their specific sequence, attracted much attention by scientists and pharmacologists who were devoted at developing the antisense approach into a new pharmacological tool.

For example, several antisense oligonucleotides have' been shown to arrest hematopoietic cell proliferation (Szczylik et al., 1991), growth (Calabretta et al.; 1941), entry into the S phase of the cell cycle (Heikhila et al., 1987), reduced survival (Reed et al., 1990) and prevent receptor mediated responses (Burch and Mahan,1991).

For efficient in vivo inhibition of gene expression using antisense oligonucleotides or analogs, the oligonucleotides or analogs must fulfill the following requirements (i) sufficient specificity in binding to the target sequence; (ii) solubility in water; (iii) stability against intra- and extracellular nucleases; (iv) capability of penetration through the cell membrane; and (v) when used to treat an organism, low toxicity.

Unmodified oligonucleotides are typically impractical for use as antisense sequences since they have short in vivo half-lives, during which they are degraded rapidly by nucleases. Furthermore, they are difficult to prepare in more than milligram quantities. In addition, Such oligonucleotides are poor cell membrane penetrators.

Thus it is apparent that in order to meet all the above listed requirements, oligonucleotide analogs need to be devised in a suitable manner. Therefore, an extensive search for modified oligonucleotides has been initiated.

For example, problems arising in connection with double-stranded DNA (dsDNA) recognition through triple helix formation have been diminished by a clever "switch back" chemical linking, whereby a sequence of polypurine on one strand is recognized, and by "switching back", a homopurine sequence on the other strand can be recognized. Also, good helix formation has been obtained by using artificial bases. thereby improving binding conditions with regard m ionic strength and pH.

<u>Oligonucleotide analogs</u>

In addition, in order to improve half-life as well as membrane penetration, a large number of variations in polynucleotide backbones have been done.

Oligonucleotides can be modified either in the base, the sugar or the phosphate

moiety. These modifications include, for example, the use of methylphosphonates, monothiophosphates, dithiophosphates, phosphoramidates, phosphate esters, bridged phosphorothioates, bridged phosphoramidates, bridged methylenephosphonates, dephospho internucleotide analogs with siloxane bridges, carbonate brides, carboxymethyl ester bridges, carbonate bridges, carboxymethyl ester bridges; acetamide bridges, carbonate bridges, thioether bridges, sulfoxy bridges, sulfono bridges, various "plastic" DNAs, .alpha.-anomeric bridges and borane derivatives.

International patent application WO 89/12060 discloses various building blocks for synthesizing oligonucleotide analogs, as well as oligonucleotide analogs formed by joining such building blocks in a defined sequence. The building blocks may be either "rigid" (i.e., containing a ring structure) or "flexible" (i.e., lacking or ring structure). In both cases, the building blocks contain a hydroxy group and a mercapto group, through which the building blocks are said to join to form oligonucleotide analogs. The linking moiety in the oligonucleotide analogs is selected from the group consisting of sulfide (-S-), sulfoxide (-SO-), and sulfone ($-SO_2-$).

International patent application WO 92/20702 describe an acyclic oligonucleotide which includes a peptide backbone on which any selected chemical nucleobases or analogs are stringed and serve an coding characters as they do in natural DNA or RNA. These new compounds, known as peptide nucleic acids (PNAs), are not only more stable in cells than their natural counterparts, but also bind natural DNA and RNA, 50 to 100 times more tightly than the natural nucleic acids cling to each other. PNA oligomers can be synthesized from the four protected monomers containing thymine, cytosine, adenine and guanine by Merrifield solid-phase peptide synthesis. In order to increase solubility in water and to prevent aggregation, a lysine amide group is placed at the C-terminal region.

Thus, in one preferred aspect antisense technology requires pairing of messenger RNA wish an oligonucleotide to form a double helix that inhibits translation. The concept of antisense-mediated gone therapy was already introduced in 1978 for cancer therapy. This approach was based on certain genes that are crucial in cell division and growth of cancer cell. Synthetic fragments of genetic substance DNA can achieve this

goal. Such molecules bind to the targeted gene molecules in RNA of tumor cells, thereby inhibiting the translation of the gates and resulting in dysfunctional growth of these cells. Other mechanisms has also been proposed. These strategies have been used, with some success is treatment of cancers, as well or other illnesses, including viral and other infectious diseases. Antisense oligonucleotides are typically synthesized in lengths of 13-30 nucleotides. The life span of oligonucleotide molecules in blood is rather shots. Thus, they have to be chemically modified to prevent destruction by ubiquitous nucleases present in the body. Phosphorothioates are very widely used modification in antisense oligonucleotide ongoing clinical trials. A new generation of antisense molecules consist of hybrid antisense oligonucleotide with a central portion of synthetic DNA while four bases on each end have been modified with 2'O-methyl ribose to resemble RNA. In preclinical studies in laboratory animals, such compounds have demonstrated greater stability to metabolism in body tissues and an improved safety profile when compared with the first-generation unmodified phosphorothioate (Hybridon Inc. news). Dozens of other nucleotide analogs have also been tested in antisense technology.

RNA oligonucleotides tray also be used for antisense inhibition as they form a stable RNA--RNA duplex with the target, suggesting efficient inhibition However, due to their low stability RNA oligonucleotides are typically expressed inside the cells using vectors designed for this purpose. This approach is favored when attempting to target a mRNA that encodes an abundant and long-lived protein.

Recent scientific publications have validated the efficacy of antisense compounds in animal models of hepatitis, cancers, coronary artery restenosis and other diseases. The first antisense drug was recently approved by the FDA. This drug Fomivirsen, developed by Isis, is indicated for local treatment of cytomegalovirus in patients with AIDS who are intolerant of or have a contraindication to other treatments for CMV retinitis or who were insufficiently responsive to previous treatments for CMV retinitis (Pharmacotherapy News Network).

Several antisense compounds are now in clinical trials in the United States. These include locally administered antivirals, systemic cancer therapeutics. Antisense

therapeutics has the potential to treat many life threatening diseases with a number of advantages over traditional drugs. Traditional drugs intervene after a disease-causing protein is formed. Antisense therapeutics, however, block mRNA transcription/translation and intervene before a protein is formed, and since antisense therapeutics target only one specific mRNA, they should be more effective with fewer side effects than current protein-inhibiting therapy.

A second option for disrupting gene expression at the level of transcription uses synthetic oligonucleotides capable of hybridizing with double stranded DNA. A triple helix is formed. Such oligonucleotides may prevent binding of transcription factors to the gene's promoter and therefore inhibit transcription. Alternatively they may prevent duplex unwinding and, therefore, transcription of genes within the triple helical structure.

Thus, according to a further aspect of the present invention there is provided a pharmaceutical composition comprising the antisense oligonucleotide described herein and a pharmaceutically acceptable carries. The pharmaceutically acceptable carrier can be, for example, a liposome loaded with the antisense oligonucleotide. Formulations for topical administration may include, but are not limited to, lotions, ointments, gels, creams, suppositories, drops, liquids, sprays and powders. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable. Compositions for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, sachets, capsules or tablets. Thickeners, diluents, flavorings, dispersing aids, emulsifiers or binders may be desirable. Formulations for parenteral administration may include but ate not limited to, sterile aqueous solutions which tray also contain buffers, diluents and other suitable additives.

According to still a further aspect of the present invention there is provided a ribozyme comprising the antisense oligonucleotide described herein and a ribozyme sequence fused thereto. Such a ribozyme is readily synthesizable using solid phase oligonucleotide synthesis.

Ribozymes are being increasingly used for the sequence-specific inhibition of gene expression by the cleavage of mRNAs encoding proteins of interest. The possibility of designing ribozymes to cleave any specific target RNA has rendered them valuable toots in both basic research and therapeutic applications. In the therapeutics area, ribozymes have been exploited to target viral RNAs in infectious diseases, dominant oncogenes in cancers and specific somatic mutations in genetic disorders. Most notably, several ribozyme gene therapy protocols for HIV patients are already in Phase 1 trials. More recently, ribozymes have been used for transgenic animal research, gene target validation and pathway elucidation Several ribozymes are in various stages of clinical trials. ANGIOZYME was the first chemically synthesized ribozyme to be studied in human clinical orals. ANGIOZYME specifically inhibits formation of VEGF-r (Vascular Endothelial Growth Factor receptor), a key component in the angiogenesis pathway, Ribozyme Pharmaceuticals, Inc., as well as other firms have demonstrated the importance of anti-angiogenesis therapeutics in animal models. HEPTAZYME, a ribozyme designed to selectively destroy Hepatitis C Virus (HCV) RNA, was found effective in decreasing Hepatitis C viral RNA in cell culture assays (Ribozyme Pharmaceuticals, Incorporated-WEB home page).

According to yet a further aspect of the present invention there is provided a recombinant or synthetic (i.e., prepared using solid phase peptide synthesis) protein comprising a polypeptide capable of binding to an ErbB receptor and which is at least 80%, preferably at least 85%, more preferably at least 90% or more, most preferably at least 95% or more or 100% identical or homologous (identical+similar) to a novel splice variant comprising the receptor binding EGF domain of an ErbB ligand with the proviso that said ligand is devoid of the C-loop of the receptor binding EGF domain.

Most preferably the polypeptide includes at least a portion of the ErbB ligand splice variants of the present invention that may include amino acids spanning cyteines 1 to 4 but are absent cysteines 5 and 6 of the receptor binding EGF domain.

Additionally or alternatively, the polypeptide according to this aspect of the present invention is preferably encoded by a polynucleotide hybridizable with SEQ ID NOs: 128 to 139 and 148 to 192, or a portion thereof under any of the stringent or moderate hybridization conditions described above for long nucleic acids. Still additionally or

alternatively, the polypeptide according to this aspect of the present invention is preferably encoded by a polynucleotide at least 80%, at least 85%, at least 90%, at least 95%, or 100%, identical with the sequences disclosed herein that encode the splice variants lacking the C-loop of the receptor binding EGF domain.

Thus, this aspect of the present invention encompasses (i) polypeptides as set forth in SEQ ID NOs: 73 to 84 and 93 to 127; (ii) fragments thereof; (iii) polypeptides homologous thereto; and (iv) altered polypeptide characterized by mutations, such as deletion, insertion or substitution of one or more amino acids, either naturally occurring or man induced, either random or in a targeted fashion, either natural, non-natural or modified at or after synthesis, with the proviso that the C-loop is absent form the receptor binding domain.

According to still a further aspect of the present invention there is provided a pharmaceutical composition comprising, as an active ingredient the recombinant protein described herein and a pharmaceutical acceptable carrier which is further described above.

<u>Peptides</u>

As used herein in the specification and in the claims section below the phrase "derived from a polypeptide" refers to peptides derived from the specified protein or proteins and further to homologous peptides derived from equivalent regions of proteins homologous to the specified proteins of the same or other species. The term further relates to permissible amino acid alterations and peptidomimetics designed based on the amino acid sequence of the specified proteins or their homologous proteins.

As used herein in the specification and in the claims section below the term "amino acid" is understood to include the 20 naturally occurring amino acids; those amino acids often modified post-translationally in vivo, including for example hydroxyproline, phosphoserine and phosphothreonine; and other unusual amino acids including, but not limited to, 2-aminoadipic acid: hydroxylysine isodesmosine, nor-valine, nor-leucine and ornithine. Furthermore, the term "amino acid" includes both D- and L-amino acids, Further elaboration of the possible amino acids usable

according to the present invention and examples of non-natural amino acids are given hereinunder.

Hydrophilic aliphatic natural amino acids can be substituted by synthetic amino acids, preferably Nleu, Nval and/or α-aminobutyric acid or by aliphatic amino acids of the general formula--$HN(CH_2)_n$ COOH, wherein n=3-5, as well as by branched derivatives thereof, wherein an alkyl group, for example, methyl, ethyl or propyl, is located at any one or more of the n carbons.

Each one, or more, of the amino acids can include a D-isomer thereof. Positively charged aliphatic carboxylic acids, such as, but not limited to, H_2 $N(CH_2)_n$ COOH, wherein n=2-4 and H_2 N--C(NH)--$NH(CH_2)_n$ COOH, wherein n=2-3, as well as by hydroxy Lysine, N-methyl Lysine or ornithine (Orn) can also be employed. Additionally, enlarged aromatic residues, such as, but not limited to, H_2 N--(C_6 H_6)--CH_2 --COOH, p-aminophenyl alanine, H_2 N--F(NH)--NH--(C_6 H_6)--CH_2 --COOH, p-guanidinophenyl alanine or pyridinoalanine (Pal) can also be employed. Side chains of amino acid derivatives (if these are Ser, Tyr, Lys, Cys or Orn) can be protected-attached to alkyl, aryl, alkyloyl or aryloyl moieties. Cyclic derivatives of amino acids can also be used. Cyclization can be obtained through amide bond formation, e.g., by incorporating Glu, Asp, Lys, Orn, di-amino butyric (Dab) acid, di-aminopropionic (Dap) acid at various positions is the chain (--CO--NH or --NH--CO bonds). Backbone to backbone cyclization can also be obtained through incorporation of modified amino acids of the formulas H--N((CH_2)$_n$ --COOH)--C(R)H--COOH or H--N((CH_2)$_n$ --COON)--C(R)H--NH_2, wherein n=1-4, and further wherein R is any natural or non-natural side chain of an amino acid. Cyclization via formation of S--S bonds through incorporation of two Cys residues is also possible. Additional side-chain to side chain cyclization can be obtained via formation of an interaction bond of the formula -(-CH_2 -)$_n$ -S-CH_2 --C--, wherein n=1 or 2, which is possible, for example, through incorporation of Cys or homoCys and reaction of its free SH group with, e.g., bromoacetylated Lys, Orn, Dab or Dap, Peptide bonds (-CO-NH-) within the peptide may be substituted by N-methylated bonds (-$N(CH_3)$- CO-), ester bonds (-C(R)H-CO--O-C(R)-N-), ketomethylene bonds (-CO-CH_2-), α-aza bonds (-NH-N(R)-CO-),

wherein R is any alkyl, e.g., methyl, carba bonds (-CH_2-NH-), hydroxyethylene bonds (-CH(OH)-CH_2-), thioamide bonds (-CS-NH-), olefinic double bonds (-CH=CH-), retro amide bonds (-NH-CO-), peptide derivatives (-N(R)-CH_2-CO-), wherein R is the "normal" side chain, naturally presented on the carbon atom. These modifications can occur at any of the bonds along the peptide chain and even at several (2-3) at the same time. Natural aromatic amino acids, Trp, Tyr and Phe, may be substituted far synthetic port-natural acid such as TIC, naphthylelanine (Nol), ring-methylated derivatives of Phe, halogenated derivatives of Phe or o-methyl Tyr.

Display libraries

According to still another aspect of the present invention there is provided a display library comprising a plurality of display vehicles (such as phages, viruses or bacteria) each displaying at least 5-10 or 15-20 consecutive amino acids derived from a polypeptide at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical or homologous (identical+similar) to SEQ ID Nos:73 to 84 and 93 to 127.

According to a preferred embodiment of this aspect of the present invention substantially every 5-10 or 15-20 consecutive amino acids derived from the polypeptide at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical or homologous (identical+similar) to SEQ ID NOs:73 to 84 and SEQ ID NOS:93 to 127 are displayed by at least one at the plurality of display vehicles, so as to provide a highly representative library. Preferably, the consecutive amino acids or amino acid analogs of the peptide or peptide analog according to this aspect of the present invention are derived from SEQ ID NOs.:73 to 84 and 93 to 127, with the proviso that these peptides are devoid of the C-loop of the EGF domain.

Methods of constructing display libraries are well known in the art, such methods are described, for example, in Young A C, et al., "The three-dimensional structures of a polysaccharide binding antibody to Cryptococcus neoformans and its complex with a peptide from a phage display library: implications for the identification of peptide mimotopes" J Mol Biol Dec 12, 1997;274(4):622-34; Giebel L B et al. "Screening of cyclic peptide phage libraries identifies ligands that bind streptavidin with high affinities" Biochemistry Nov 28, 1995;34 (47):15430-5; Davies E L et al., "Selection

of specific phage-display antibodies using libraries derived from chicken immunoglobulin genes" J Immunol Methods Oct. 12, 1995;186(1):125-35; Jones C et al. "Current trends in molecular recognition and bioseparation" J Chromatogr A Jul 14, 1995;707(1):3-22; Deng S J et al. "Basis for selection of improved carbohydrate-binding single-chain antibodies from synthetic gene libraries" Proc Natl Acad Sci U S A May 23, 1995;92(11):4992-6; and Deng S J et al. "Selection of antibody single-chain variable fragments with improved carbohydrate bidding by phage display" J Biol Chem Apr 1, 1994;269(13):9533-8, which are incorporated herein by reference. Display libraries according to this aspect of the present invention can be used to identify and isolate polypeptides which are capable of up- or down-regulating ErbB activity.

Antibodies

According to still another aspect of the present invention there is provided an antibody comprising at least the antigen binding portion of an immunoglobulin specifically recognizing and binding a polypeptide at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical or homologous (identical+similar) to SEQ ID NOs: 73 to 84 and 93 to 127 with the proviso that these antibodies do not bind significantly to the C-loop of an intact EGF domain.

The present invention can utilize serum immunoglobulins, polyclonal antibodies or fragments thereof, (i.e., immunoreactive derivative of an antibody), or monoclonal antibodies or fragments thereof. Monoclonal antibodies of purified fragments of the monoclonal antibodies having at least a portion of an antigen bidding region, including such as Fv, F(abl)2, Fab fragments (Harlow and Lane, 1988 Antibody, Cold Spring Harbor); single chain antibodies (U.S. Pat. No. 4,946,778), chimeric or humanized antibodies and complementarily determining regions (CDR) may be prepared by conventional procedures. Purification of these serum immunoglobulins antibodies or fragments can be accomplished by a variety of methods known to those of skill including, precipitation by ammonium sulfate or sodium sulfate followed by dialysis against saline, ion exchange chromatography, affinity or immunoaffinity chromatography as well as gel filtration, zone electrophoresis, etc. (see Goding in, Monoclonal Antibodies: Principles and Practice, 2nd ed., pp. 104-126, 1986, Orlando,

Fla., Academic Press). Under normal physiological conditions antibodies are found in plasma and other body fluids and in the membrane of certain cells and are produced by lymphocytes of the type denoted B cells or their functional equivalent. Antibodies of the IgG class are made up of four polypeptide chains linked together by disulfide bonds. The four chains of intact IgG molecules are two identical heavy chains referred to as H-chains and two identical light chains referred to as L-chains. Additional classes includes IgD, IgE, IgA, IgM and related proteins.

Monoclonal antibodies

Methods for the generation and selection of monoclonal antibodies are well known in the art, as summarized for example in reviews such as Tramontano and Schloeder, Methods in Enzymology 178, 551-568, 1989. A recombinant or synthetic ErbB ligand or a portion thereof of the present invention may be used to generate antibodies in vitro. More preferably, the recombinant or synthetic ErbB ligand of the present invention is used to elicit antibodies in vivo. In general, a suitable host animal is immunized with the recombinant or synthetic ErbB ligand of the present invention or a portion thereof including at least one continuous or discontinuous epitope. Advantageously, the animal host used is a mouse of an inbred strain. Animals are typically immunized with a mixture comprising a solution of the recombinant or synthetic ErbB ligand of the present invention or portion thereof in a physiologically acceptable vehicle, and any suitable adjuvant, which achieves as enhanced immune response to the immunogen. By way of example, the primary immunization conveniently may be accomplished with a mixture of a solution of the recombinant or synthetic ErbB ligand of the present invention or a portion thereof and Freund's complete adjuvant, said mixture being prepared in the form of a water-in-oil emulsion. Typically the immunization may be administered to the animals intramuscularly, intradermally, subcutaneously, intraperitoneally, into the footpads, or by any appropriate route of administration. The immunization schedule of the immunogen may be adapted as required, but customarily involves several subsequent or secondary immunizations using a milder adjuvant such as Freund's incomplete adjuvant. Antibody titers and specificity of binding can be determined during the immunization schedule by any convenient method including by way of example

radioimmunoassay, or enzyme linked immunosorbant assay, which is known as the ELISA assay. When suitable antibody titers are achieved, antibody producing lymphocytes from the immunized animals are obtained, and these are cultured, selected and closed, as is known in the art. Typically, lymphocytes may be obtained in large numbers from the spleens of immunized animals, but they may also be retrieved from the circulation, the lymph nodes or other lymphoid organs. Lymphocyte are then fused with any suitable myeloma cell line, to yield hybridomas, as is well known in the art. Alternatively, lymphocytes may also be stimulated to grow in culture; and may be immortalized by methods known in the art including the exposure of these lymphocytes to a virus; a chemical or a nucleic acid such as an oncogene, according to established protocols. After fusion, the hybridomas ate cultured under suitable culture conditions, for example in multiwell plates, and the culture supernatants are screened to identify cultures containing antibodies that recognize the hapten of choice. Hybridomas that secrete antibodies that recognize the recombinant or synthetic NRG-4 of the present invention are cloned by limiting dilution and expanded, under appropriate culture conditions. Monoclonal antibodies are purified and characterized in terms of immunoglobulin type and binding affinity.

Pharmaceutical compositions for Regulation of ErbB receptor activity

Thus, according to yet another aspect of the present invention there is provided a pharmaceutical composition comprising, as an active ingredient, an agent for regulating an ErbB receptor mediated activity in vivo or in vitro. The following embodiments of the present invention are directed at intervention with ErbB ligand activity and therefore with ErbB receptor signaling.

According to yet another aspect of the present invention there is provided a method of regulating an endogenous protein affecting ErbB receptor activity in vivo or in vitro. The method according to this aspect of the present invention is effected by administering an agent for regulating the endogenous protein activity in vivo, the endogenous protein being at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical or homologous (identical+similar) to SEQ ID NOs: 73 to 84 and 93 to 127, with the proviso that it is devoid of the C-loop of the intact EGF domain.

An agent which can be used according to the present invention to upregulate the activity of the endogenous protein can include, for example, an expressible sense polynucleotide at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical with SEQ ID NOs:128 to 139 and 148 to 192, with the proviso that it does not encode the C-loop of the intact EGF domain.

An agent which can be used according to the present invention to downregulate the activity of the endogenous protein can include, for example, an expressible antisense polynucleotide at least 80%, at least 85%, at least 90%, at least 95%, or 100%, identical with a portion of SEQ ID Nos:128 to 139 and 148 to 192, with the proviso that it does not encode the C-loop of the intact EGF domain.

Alternatively, an agent which can be used according to the present invention to downregulate the activity of the endogenous protein can include, for example, an antisense oligonucleotide or ribozyme which includes a polynucleotide or a polynucleotide analog of at least 10 bases, preferably between 10 and 15, more preferably between 15 and 20 bases, most preferably, at least 17-40 bases which is hybridizable in vivo, under physiological conditions, with a portion of a polynucleotide strand encoding a polypeptide at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical or homologous (identical+similar) to SEQ ID NOs:128 to 139 and 148 to 192.

[Do you have a standard clause on siRNA technology? This seems to be the new preferred methodology of gene silencing.]

Still alternatively, an agent which can be used according to the present invention to downregulate the activity of the endogenous protein can include, for example, an peptide or a peptide analog representing a stretch of at least 6-10, 10-15, or 15-20 consecutive amino acids or analogs thereof derived from a polypeptide at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical or homologous (identical+similar) to SEQ ID NOs: 73 to 84 and 93 to 127.

Peptides or peptide analogs containing the interacting EGF-like domain according to the present invention will compete by protein interactions to form protein complexes with ErbB receptor, inhibiting or accelerating the pathways in which ErbB

ligands are involved.

The following biochemical and molecular systems are known for the characterization and identification of protein-protein interaction and peptides as substrates, through peptide analysis, which systems can be used to identify inhibitory peptide sequences. One such system employs introduction of a genetic material encoding a functional protein or a mutated form of the protein, including amino acid deletions and substitutions, into cells. This system, can be used to identify functional domains of the protein by the analysis of its activity and the activity of its derived mutants in the cells. Another such system employs the introduction of small encoding fragments of a gene into cells, e.g., by means of a display library or a directional randomly primed cDNA library comprising fragments of the gene, and analyzing the activity of the endogenous protein in their presence (see, for example, Gudkov et al. (1993) "Isolation of genetic suppressor elements, including resistance to topoisomerase II interactive cytotoxic drugs, from human topoisomerase II cDNA" Proc. Natl. Acad. Sci. USA 90:3231-3236; Gudkov and Robinson (1997) "Isolation of genetic suppressor elements (GSEs) from random fragment cDNA libraries in retroviral vectors" Methods Mol Biol 69;221-240; and Pestov et al. (1999) "Flow Cytometric Analysis of the cell cycle in transfected cells without cell fixation" Bio Techniques 26:102-106). Yet an additional system is realized by screening expression libraries with peptide domains, as exemplified, for example, by Yamabhai et al. (1998 "Intersectin, a Novel Adaptor Protein with Two Eps15 Homology and Five Src Homology 3 Domains". J Biol Chem 273: 31401-31407). In yet another such system overlapping synthetic peptides derived from specific gene products are used to study and affect in vivo and in vitro protein-protein interactions. For example, synthetic overlapping peptides derived from the HIV-1 gene (20-30 amino acids) were assayed for different viral activities (Baraz et al. (1998) "Human immunodeficiency virus type 1 Vif derived peptides inhibit the viral protease and arrest virus production" FEBS Letters 441:419-426) and were found to inhibit purified viral protease activity; bind to the viral protease; inhibit the Gag-Pol polyprotein cleavage; and inhibit mature virus production in human cells.

EXAMPLES

Materials and Methods.

EST, genomic and non redundant databases were searched for homology particularly to the EGF-like domains of various ErbB ligands by BLAST and Smith-Waterman based searches (Altschul *et al.*, 1997; Samuel and Altschul, 1990; Smith and Waterman, 1981). BLASTN, BLASTP and TBLASTN – based searches were performed using the National Center for Biological Information (NCBI) node, utilizing both the search engines and databases offered at this site. Multiple sequence alignments were performed using ClustalX (Version 1.81 for Windows); (Chenna et. al. 2003) . Smith-Waterman based searches were performed using a software package and Compugen Biocellerator maintained at the European Molecular Biology Laboratory (EMBL-interface). Profile-based searches were also performed using this Biocellerator; Sequence profiles were generated from ClustalX multiple sequence alignments of proteins using the software PROFILEWEIGHT, which is provided as a software component of the EMBL-interface Compugen Biocellerator. Profile searches were then performed against DNA databases, using the program TPROFILESEARCH (Compugen Biocellerator at EMBL; program version 1.9) . The databases scanned for the Biocellerator searches were in this case maintained at the EMBL site.

Sequences of defined names or accession numbers were retrieved directly using the NCBI Entrez sequence retrieval tools. DNA sequence translations were performed using the program Transeq, a component of the EMBOSS package and provided by the EMBL-European Bioinformatics Institute Node (Rice et. al.; Trends Genet. 2000 Jun;16(6):276-7). Domain architecture was defined with the aid of reading the literature and also by use of the SMART (Simple Modular Architecture Research Tool; EMBL) (Letunic et. al.; Nucleic Acids Res. 2002 Jan 1;30(1):242-4). Default settings were used with the use of all bioinformatics tools, unless otherwise indicated in the text. At the time of the writing of this manuscript the above programs and Web interfaces could be accessed from the sites shown in Table 4.

Table 4: Resources/tools used for bioinformatics analyses

Name	Site

Entrez Server	http://www.ncbi.nlm.nih.gov/Entrez/
Blast Server	http://www.ncbi.nlm.nih.gov/blast/
Compugen Biocellerator Server (EMBL)	http://eta.embl-heidelberg.de:8000/misc/
Compugen PROFILEWEIGHT	http://eta.embl-heidelberg.de:8000/profw/
Emboss Transeq Server	http://www.ebi.ac.uk/emboss/transeq/
SMART Server	http://smart.embl-heidelberg.de/
ClustalX	ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/

Typical members of the ErbB ligand family have already been described elsewhere (Harari *et al.*, 1999; Harris *et al.*, 2003; Strachan *et al.*, 2001). Protein sequences for these ligands were extracted from the NCBI server by utilization of the Entrez sequence retrieval tool as well as by BLASTP searches against the NR protein database. Subsequently corresponding cDNA sequences were pulled out as reference links to the protein sequences, or by TBLASTN searches against the NR DNA database. Finally, genomic contigs encoding at least protions of the ErbB ligands were extracted by performing TBLASTN searches against the NCBI human and mouse genomic databases. Accession numbers of representative sequences are provided in Table 5. It should be noted, that these sequences are often redundantly represented in the database, and furthermore, there are the existence of alternative splice variants for some ligands. Thus the accession numbers given here are representative ones only. Reference to alternative accession numbers may be incorporated into the text.

Table 5. Accession numbers pertaining to genomic, transcript and protein sequences encoding different ErbB-ligands

GENE	NCBI accession # cDNA	NCBI accession # Protein	NCBI accession # Genomic Contig
NRG 1 Alpha	AF491780 *	AM71141.1	NT_007995.10
NRG1 Beta	AF491780 *	AAM71136.1	
NRG2 Alpha	NP_004874	NM_013982	NT_029289
NRG2 Beta	NM_013983	NP_053586.1	
NRG3	XM_170640.1	P56975	NT_033890.2
NRG4	NM_138573.1	NP_612640.1	NT_024654.12
EGF	NM_001963.2	NP_001954.1	NT_028147.9
TGF alpha	K03222	P01135	NT_022184.9
Amphiregulin	M30704	AAA51781.1	NT_006216.11
HB-EGF	BC033097	AAH33097.1	NT_034777.1
Betacellulin	S55606	P35070	NT_034698.1
Epiregulin	NM_001432	NP_001423.1	NT_006216.11
Epigen (Mouse)	AJ291391	CAC39435.1	NT_039307.1
Epigen (Human)			NT_006216.1
Lin-3 (C. elegans)	NM_171919	NP_741490	
Argos (Dros. melanogaster)	NM_079383	NP_524107.2	AE003527
Argos (Musca domestica)	AF038405	AAB92420	
Argos (Dros virilis)	AB089249	BAC56702	

* Numerous NRG1 variants are provided with this single accession.

The present invention has been described with reference to specific preferred embodiments and examples. It will be appreciated by the skilled artisan that many possible alternatives will be apparent within the scope of the present invention which is not intended to be limited by the specific embodiments exemplified herein but rather by the following claims.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-402, 1997.

Barbacci, E.G.; Guarino, B.C.; Stroh, J.G.; Singleton, D.H.; Rosnack, K.J.; Moyer, J.D.; and Andrews, G.C. The structural basis for the specificity of epidermal growth factor and heregulin binding. *J Biol Chem* , 270(16):9585-9589, 1995.

Campion, S.R., and Niyogi, S.K. Interaction of epidermal growth factor with its receptor. *Prog Nucleic Acid Res Mol Biol* , 49:353-383, 1994.

Carpenter, G., and Cohen, S. Epidermal growth factor. *J. Biol. Chem.*, 265:7709-7712, 1990.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. and Thompson, J. D. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31, 3497-3500, 2003.

Crovello, C.S.; Lai, C.; Cantley, L.C.; and Carraway, K.L., 3rd. Differential signaling by the epidermal growth factor-like growth factors neuregulin-1 and neuregulin-2. *J Biol Chem* , 273(41):26954-26961, 1998.

Darnell, J.; Lodish, H.; and Baltimore, D., Molecular Cell Biology, Scientific American Books, USA (1986)

Defeo-Jones, D.; Tai, J.Y.; Vuocolo, G.A.; Wegrzyn, R.J.; Schofield, T.L.; Riemen, M.W.; and Oliff, A. Substitution of lysine for arginine at position 42 of human transforming growth factor-a eliminates biological activity without changing internal disulfide bonds. *Mol. Cell. Biol.*, 9:4083-4086, 1989.

Engler, D.A.; Campion, S.R.; Hanser, M.R.; Cook, J.S.; and Niyogi, S.K. Critical functional requirements for the guanidinium group of the arginine 41 side chain of human epidermal growth factor as revealed by mutagenic inactivation and chemical reactivation. *J. Biol. Chem.* , 267:2274-2281, 1992.

Falls, D.L. Neuregulins: functions, forms, and signaling strategies. *Exp Cell Res.*, 284(1):14-30, 2003.

Groenen, L.C.; Nice, E.C.; and Burgess, A.W. Structure-function relationships for the EGF/TGF-a family of mitogens. *Growth Factors*, 11:235-257, 1994.

Harari, D.; Tzahar, E.; Romano, J.; Shelly, M.; Pierce, J.H.; Andrews, G.C.; and Yarden, Y. Neuregulin-4: a novel growth factor that acts through the ErbB-4 receptor tyrosine kinase. *Oncogene*, 18(17):2681-2689, 1999.

Harari, D., and Yarden, Y. Molecular mechanisms underlying ErbB2/HER2 action in breast cancer. *Oncogene*, 19(53):6102-6114, 2000.

Harris, R.C.; Chung, E.; and Coffey, R.J. EGF receptor ligands. *Experimental Cell Research*, 284:2-13, 2003.

Howes, R.; Wasserman, J.D.; and Freeman, M. In vivo analysis of Argos structure-function. Sequence requirements for inhibition of the Drosophila epidermal growth factor receptor. *Journal of Biological Chemistry*, 273(7):4275-4281, 1998.

Jin, M.H.; Sawamoto, K.; Ito, M.; and Okano, H. The interaction between the Drosophila secreted protein argos and the epidermal growth factor receptor inhibits dimerization of the receptor and binding of secreted spitz to the receptor. *Mol Cell Biol*, 20(6):2098-2107, 2000.

Jones, J.T.; Akita, R.W.; and Sliwkowski, M.X. Binding specificities and affinities of egf domains for ErbB receptors. *FEBS Lett*, 447(2-3):227-231, 1999.

Jorissen, R.N.; Walker, F.; Pouliot, N.; Garrett, T.P.; Ward, C.W.; and Burgess, A.W. Epidermal growth factor receptor: mechanisms of activation and signalling. *Exp Cell Res*, 284(1):31-53, 2003.

Kurreck, J. Antisense technologies. Improvement through novel chemical modifications. *Eur J Biochem*, 270(8):1628-1644, 2003.

Maniatis, T.; Fritsch, E.F.; and Sambrook, J. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor (New York: Cold Spring Harbor Laboratory), 1982.

Moghal, N., and Sternberg, P.W. The epidermal growth factor system in Caenorhabditis elegans. *Exp Cell Res*, 284(1):150-159, 2003.

Pennock, S.; and Wang, Z. Stimulation of Cell Proliferation by Endosomal Epidermal Growth Factor Receptor As Revealed through Two Distinct Phases of Signaling. Mol. Cell. Biol., 23(16): 5803-15, 2003

Sarup, J.C.; Johnson, R.M.; King, K.L.; Fendly, B.M.; Lipari, M.T.; Napier, M.A.; Ullrich, A.; and Shepard, H.M. Characterization of an anti-p185HER2 monoclonal antibody that stimulates receptor function and inhibits tumor cell growth. *Growth Regul*, 1(2):72-82, 1991.

Schnepp, B.; Donaldson, T.; Grumbling, G.; Ostrowski, S.; Schweitzer, R.; Shilo, B.Z.; and Simcox, A. EGF domain swap converts a drosophila EGF receptor activator into an inhibitor. *Genes Dev*, 12(7):908-913, 1998.

Shilo, B.Z. Signaling by the Drosophila epidermal growth factor receptor pathway during development. *Exp Cell Res*, 284(1):140-149, 2003.

Strachan, L.; Murison, J.G.; Prestidge, R.L.; Sleeman, M.A.; Watson, J.D.; and Kumble, K.D. Cloning and biological activity of epigen, a novel member of the epidermal growth factor superfamily. *J Biol Chem*, 276(21):18265-18271, 2001.

Summerfield, A.E.; Hudnall, A.K.; Lukas, T.J.; Guyer, C.A.; and Staros, J.V.

Identification of residues of the epidermal growth factor receptor proximal to residue 45 of bound epidermal growth factor. *J. Biol. Chem.*, 271:19656-19659, 1996.

Tzahar, E.; Moyer, J.D.; Waterman, H.; Barbacci, E.G.; Bao, J.; Levkowitz, G.; Shelly, M.; Strano, S.; Pinkas-Kramarski, R.; Pierce, J.H.; Andrews, G.C.; and Yarden, Y. Pathogenic poxviruses reveal viral strategies to exploit the ErbB signaling network. *Embo J*, 17(20):5948-5963, 1998.

Vinos, J., and Freeman, M. Evidence that Argos is an antagonistic ligand of the EGF receptor. *Oncogene*, 19(31):3560-3562, 2000.

Yarden, Y., and Sliwkowski, M.X. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol*, 2:127-137, 2001.

CLAIMS

1. A polypeptide comprising a splice variant of an ErbB ligand encoded by differential exon usage comprising a truncated ErbB-Receptor-binding EGF domain devoid of the C-loop of the EGF domain.
2. The polypeptide according to claim 1 wherein the splice variant comprises a truncated receptor binding EGF domain comprising only the first four of the six conserved cysteines found in an intact EGF domain.
3. The polypeptide of claim 2 wherein the fourth conserved cysteine of the truncated ErbB-Receptor binding EGF domain is the penultimate amino acid at the C terminus of the polypeptide.
4. The polypeptide according to claim 3 comprising the sequence of any one of SEQ ID NOS:73 to 84.
5. The polypeptide according to claim 3 having the sequence of any one of SEQ ID NOS:93-110.
6. The polypeptide according to claim 2 wherein the splice variant comprises a receptor-binding EGF domain having only the first four of the six conserved cysteines found in an intact EGF domain, further comprising an amino acid sequence encoded by an alternative exon other than the second exon encoding conserved cysteines five and six the of the intact ErbB receptor-binding EGF domain.
7. The polypeptide according to claim 6 having the sequence of any one of SEQ ID NOS:111-127.
8. The polypeptide according to claim 2 wherein the splice variant comprises a receptor binding EGF domain having only the first four of the six conserved cysteines found in an intact EGF domain, wherein the splice variant has at least 90% homology to the aligned amino acid sequence of the same fragment in the EGF domain of a known ErbB ligand between cysteine 1 and cysteine 4.
9. The polypeptide of claim 8 wherein the splice variant has at least 95% homology to the aligned amino acid sequence of the same fragment in the EGF domain of a known ErbB ligand between cysteine 1 and cysteine 4.
10. The polypeptide of claim 7 or claim 8 wherein the N terminal flanking sequences preceding the cysteine 1 are at least 90% homologous to the same

sequence in the EGF domain of a known ErbB ligand.

11. The polypeptide of any one of claims 1 to 9 wherein the splice variant retains binding affinity to at least one member of the ErbB/EGF receptor family.
12. The polypeptide of claim 10 which retains binding affinity to the receptor on cells with significantly reduced biological activity compared to an equimolar amount at least one known agonist ligand.
13. An isolated polynucleotide encoding a splice variant of an ErbB ligand comprising a truncated ErbB-Receptor-binding EGF domain devoid of the C-loop of the EGF domain.
14. The polynucleotide according to claim 13 wherein the splice variant comprises a truncated receptor-binding EGF domain comprising only the first four of the six conserved cysteines found in an intact EGF domain.
15. The polynucleotide of claim 14 wherein the fourth conserved cysteine of the encoded truncated ErbB-Receptor binding EGF domain is the penultimate amino acid at the C terminus of the polypeptide.
16. The polynucleotide according to claim 15 comprising the sequence of any one of SEQ ID NOS:128 to 139.
17. The polynucleotide according to claim 15 having the sequence of any one of SEQ ID NOS:148 to 165.
18. The polynucleotide according to claim 14 wherein the encoded splice variant comprises a receptor-binding EGF domain having only the first four of the six conserved cysteines found in an intact EGF domain, further comprising an amino acid sequence encoded by an alternative exon other than the second exon encoding conserved cysteines five and six the of the intact ErbB receptor-binding EGF domain.
19. The polynucleotide according to claim 18 having the sequence of any one of SEQ ID NOS:166-182.
20. The polynucleotide according to claim 14 wherein the splice variant comprises a receptor binding EGF domain comprising only the first four of the six conserved cysteines found in an intact EGF domain, wherein the splice variant has at least 90% homology to the aligned amino acid sequence of the same fragment in the EGF domain of a known ErbB ligand between cysteine 1 and cysteine 4.

21. The polynucleotide of claim 19 wherein there is at least 95% homology to the aligned amino acid sequence of the same fragment in the EGF domain of a known ErbB ligand between cysteine 1 and cysteine 4.

22. The polynucleotide of claim 19 or claim 20 wherein the encoded N terminal flanking sequences preceding the cysteine 1 are at least 90% homologous to the same sequence in the EGF domain of a known ErbB ligand.

23. The polynucleotide of any one of claims 13 to 21 wherein the splice variant retains binding affinity to at least one member of the ErbB/EGF receptor family.

24. The polynucleotide of claim 23 which encodes a polypeptide that retains binding affinity to the receptor on cells with significantly reduced biological activity compared to an equimolar amount at least one known agonist ligand.

25. An antisense oligonucleotide capable of specifically inhibiting the expression of a polypeptide according to any one of claims 1-12.

26. A polynucleotide construct comprising an isolated polynucleotide encoding the splice variants of any one of claims 1-12.

27. A vector comprising the isolated polynucleotide encoding the splice variants of any one of claims 1-12.

28. A host cell transformed with a polynucleotide encoding the splice variants of any one of claims 1-12.

29. A host cell transformed with a polynucleotide according to any one of claims 13-24.

30. A pharmaceutical composition comprising as an active ingredient a polypeptide according to anyone of claims 1-12.

31. A pharmaceutical composition comprising as an active ingredient a polynucleotide according to anyone of claims 13-24.

32. A pharmaceutical composition comprising as an active ingredient a antisense oligonucleotide according to claim 25.

33. A method of treating a disease or disorder related to an ErbB receptor in an individual in need thereof comprising administering to the individual a therapeutically effective amount of a polypeptide according to any one of claims 1-12.

34. The method of claim 33 wherein the disease or disorder is selected

from a neoplastic disease, a hyperproliferative disease, angiogenesis, restenosis, wound healing, psychiatric disorders, neurological disorders and neurological injuries.

35. A method of treating a disease related to pathological activity of at least one ErbB receptor comprising administering a therapeutically effective amount of a polynucleotide according to any one of claims 13-24.

36. The method of claim 35 wherein the disease or disorder is selected from a neoplastic disease, a hyperproliferative disease, angiogenesis, restenosis, wound healing, psychiatric disorders, neurological disorders or neural injury.

37. A method for selectively enhancing or promoting the proliferation or differentiation of stem cells expressing ErbB receptors, comprising exposing the stem cells to an ErbB ligand splice variant, according to any one of claims 1-12.

38. The method of claim 37 wherein the stem cells are of neural, cardiac or pancreatic lineages.

ABSTRACT

The present invention relates to nucleic acid and amino acid sequences of previously unknown ErbB ligands that are splice variants of previously known ErbB ligands, to compositions comprising these sequences and uses thereof in the diagnosis, treatment, and prevention of diseases and disorders mediated by ErbB receptors. Specifically, the present invention relates to splice variants lacking the C-loop of an intact EGF receptor-binding domain.

```
Antagonists
Argos          PGYRYLFACSP-LTRLRCQRKQPCKLFTVRKRQEFLDEVNINSLCQCPK---CHRCPSHHTQSG-   Drosophila Melanogasta

Agonists
Lin-3          IEKLKEAKCKD-----YCHHNATCHVEVIFRED---RVSAVVPSCHCPQGWECTRCDRHYVQAF-   C. elegans

Vein           DRSASGIPCN--FD--YCFHNGTCRMIPDIN--------EVY--CRCPTEYFCNRCENKWPDSR-   D. Melanogasta
Gurken         ETEIQMLPCSEAYNTSFCLNGGHCFQHPMVNN-------TVFHSCLCVNDYDCERCAYKSWNGD-   D. Melanogasta
Spitz          NITFPTYKCPETFDAWYCLNDAHCFAVKIADL-------PVYS-CECAIGFMCQRCEYKEIDNT-   D. Melanogasta
Keren          NVTFPIFACPPTYVAWYCLNDGTCFTVKIHNE-------ILYN-CECALGFMCPRCEYKEIDGS-   D. Melanogasta

NRG1_alpha     TGTSHLVKCAEKEKT-FCVNGGECFMVKDLS-------NPSRYLCKCQPGFTCARCTENVPMKV-   Homo sapiens
NRG1_beta      TGTSHLVKCAEKEKT-FCVNGGECFMVKDLS-------NPSRYLCKCPNEFTCDRCQNYVMASF-   Homo sapiens
NRG2_alpha     SWSGHARKCNETAKS-YCVNGGVCYYIEGIN-------QLS---CKCPNGFFCQRCLEKLPLRL-   Homo sapiens
NRG2_beta      SWSGHARKCNETAKS-YCVNGGVCYYIEGIN-------QLS---CKCPVGYTCDRCQQFAMVNF-   Homo sapiens
NRG3           ERSEHFKPCRDKDLA-YCLNDGECFVIETLTG------SHK--HCRCKEGYQCVRCDQ-FLPKTD   Homo sapiens
NRG4           MPTDHEEPCGPSHKS-FCLNGGLCYVIPTIP-------SP---FCRCVENYTCARCEEVFLPGS-   Homo sapiens

EGF            SVRNSDSECPLSHDG-YCLHDGVCMYIEALD-------KYA---CNCVVGYICERCQYRDLKWW-   Homo sapiens
TGF_alpha      AVVSHFNDCPDSHTQ-FCFH-GTCRFLVQED-------KPA---CVCHSGYVCARCEHADLLAV-   Homo sapiens
Betacellulin   KRKGHFSRCPKQYKH-YCIK-GRCRFVVAEQ-------TPS---CVCDEGYICARCERVDLFYL-   Homo sapiens
Amphiregulin   RNRKKKNPCNAEFQN-FCIH-GECKYIEHLE-------AVT---CKCQQEYFCERCGEKSMKTH-   Homo sapiens
HB-EGF         GLGKKRDPCLRKYKD-FCIH-GECKYVKELR-------APS---CICHPGYHCERCHGLSLPVE-   Homo sapiens
Epiregulin     VAQVSITKCSSDMNG-YCLH-GQCIYLVDMS-------QNY---CRCEVGYTCVRCEHFFLTVH-   Homo sapiens
Epigen         VALKFSHPCLEDHNS-YCIN-GACAFHHELK-------QAI---CRCFTGYTCQRCEHLTLTSY-   Mus musculus

Conserved cysteine #:   1         2     3                  4 5       6
```

FIGURE 1

```
Dros._melanogaster   ----------MPTTLMLLPCMLLLLLTAAAVAVGGTRLPLEVFEITPTTS--TADKHKSL
Dros._virilis        MASIRAHSLLLLLRLMLLPLPLLLLLLMLTGGAQSTRLPLEVYELTPTASADSDAKHKSL
Musca_domestica      ---------------MLSLTIFMLATHIINACYSTRLPLEVYELTPNAAGGTDLKHKNL
                                    :*.  :::*      .  .*******:*:**.::  :  ***.*

Dros._melanogaster   QYTVVYDAK-----DISG-------AAAATG-----VASSTVKPATEQLTVVSISSTAAA
Dros._virilis        EY-AIYDPK-----ELTGAPKAAAAAAAATT-----TTSSTARPSSEKPLAIAVVSISAE
Musca_domestica      EYSTINGGSGQHFLAINGRSKQQHVSAAMEEPELKMLSSHDSKAAATKTLTVSSMGTPSA
                     :* .: . .        :.*       :**         :*     :.:: :   .::  . .:

Dros._melanogaster   EKDLAES-----------------RRHARQMLQKQQQ-------------HRSIIGG---K
Dros._virilis        QQQQQQSELEPATQ---------AGRRARQMLQQQHR-------------LSSSSSSSSNK
Musca_domestica      AATTTTSSTSTATATATTNQLDRRRSRQMLDIMQKNHHDQTGNHKLPPVLSSGEGAATG
                          *                    *::****:   ::                *     .

Dros._melanogaster   HGD--RDVRILYQVGDSEEDLPVCAPNAVCSKIDLYETPWIERQCRCPESNRMPNNVIIH
Dros._virilis        HAHSVKDLRILYQVGDSEADLPVCAPNAVCSKIDLYETPWIERQCRCPESNRMPNNVIVH
Musca_domestica      ASHSQKDVRILYQVGNSEDDLPICAPNAVCSKIDLYETPWIERQCRCPAVNRSPEIIIHH
                      ..  :*:*******:** ***:************************  ** *: :* *
                                           <----------------------->
                                                   A1 domain

Dros._melanogaster   HHSHSSGSVDS--LKYRNYYEREKMMQHKR---MLLGEF---QDKKFESLHMKKLMQKLG
Dros._virilis        HHEHPHGTMSEG-QKYRSYYEKEKLLQHKR---LLL-------DKKYESLHLKKLMQKLG
Musca_domestica      HHKETASHSNHNSEKYHTFYEHSKLAHQQQNKHLLLDAASFVGDKKFDNLHLKKLMHKLG
                     **...  .   .     **:.:**:.*: ::::    :**        ***::.**:****:***

Dros._melanogaster   AVYEDDL------DHLDQS----PDYNDALP--YAEVQDNEFP-----------RGSAHM
Dros._virilis        AVYEDDLQLPSAGDYVERS----PDYNEALPPAYEELADNELPQ--------APARSATHM
Musca_domestica      AVYEDDLNLPSDYHRHEETNSALDDSNEATLYYADEIKDNEFPAHFAMKRQHLYSNTPHM
                     *******       .   :.:       * *:*      *: ***:*            .:.**

Dros._melanogaster   RHSGHRG-SKEPATTFIGGCPSSLGVEDGHTIADKTRHYKMCQPVHKLPVCTHFRDYTWT
Dros._virilis        RHSGHRG-LKE-AVSFIGGCPSNLGVEDGHTIADKTRHYKLCQPVHKLPVCKHFRDYTWT
Musca_domestica      RHSGHTGGGHGGKISYIGGCPSGLGIEDGHTIADKTRHYKMCQPVHRLPVCRHFRDYTWT
                     ***** *  :      ::******.**:**************:*****:****  ********
                                                              <------------------
                                                                   A2 domain

Dros._melanogaster   LTTAAELNVTEQIVHCRCPRNSVTYLTKREPIGNGSPGYRYLFACSPLTRLRCQRKQPCK
Dros._virilis        LTTAAELNVTEQVVHCRCPKNSVTYLAKREPVPNSSTAYRYLFACSPLTRLRCQRKQPCK
Musca_domestica      LTTSPEMNTTEQIVHCRCPKNSVTYLTKREPSEDGNGGYKYLFACSPLTRFRCQRKQPCK
                     ***:.*:*.***:******:******:****   :..  .*:**********:*********
                     ------------------>                         <---------------
                       A2 (continued)                                 EGF domain

Dros._melanogaster   LFTVRKRQEFLDEVNINSLCQCPKGHRCPSHHTQSGVIAGESFLEDNIQTYSGYCMAND
Dros._virilis        LFTVRKRQEFLDEVNINSLCQCPKGHRCPSHHTQSGVIAGESFLEDNIQTYSGYCMAND
Musca_domestica      LFTVRKRQEFIDEVNINALCQCPKGHHCPSHHTQSGVIAGETFLEDNIQTYSGYCMVND
                     **********:******:******:*:*************:**************.**
                     ----------------------------->
                        EGF domain (continued)
```

FIGURE 2

```
                                                                                Sequence ID #

NRG1_alpha      TGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCKCQPGFTGARCTENVPMKV     1
NRG1_beta       TGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCKCPNEFTGDRCQNYVMASF     2
NRG2_alpha      SWSGHARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPNGFFGQRCLEKLPLRL     3
NRG2_beta       SWSGHARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPVGYTGDRCQQFAMVNF     4
NRG3            ERSEHFKPCRDKDLAYCLNDGECFVIETLTGSHK-HCRCKEGYQGVRCDQFLPKTD     5
NRG4            MPTDHEEPCGPSHKSFCLNGGLCYVIPTIP--SP-FCRCVENYTGARCEEVFLPGS     6

EGF             SVRNSDSECPLSHDGYCLHDGVCMYIEALDKYA---CNCVVGYIGERCQYRDLKWW     7
TGF_alpha       AVVSHFNDCPDSHTQFCFH-GTCRFLVQEDKPA---CVCHSGYVGARCEHADLLAV     8
Betacellulin    KRKGHFSRCPKQYKHYCIK-GRCRFVVAEQTPS---CVCDEGYIGARCERVDLFYL     9
Amphiregulin    RNRKKKNPCNAEFQNFCIH-GECKYIEHLEAVT---CKCQQEYFGERCGEKSMKTH     10
HB-EGF          GLGKKRDPCLRKYKDFCIH-GECKYVKELRAPS---CICHPGYHGERCHGLSLPVE     11
Epiregulin      VAQVSITKCSSDMNGYCLH-GQCIYLVDMSQNY---CRCEVGYTGVRCEHFFLTVH     12
Epigen          VALKFSHPCLEDHNSYCIN-GACAFHHELKQAI---CRCFTGYTGQRCEHLTLTSY     13
                        *      :*.: * *             * *   : * **
```

FIGURE 3

HUMAN	<--------------- Exon A --------------><----- Exon B ----->	Sequenc ID #
NRG1_alpha	....TGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCKCQPGFTGARCTENVPMKVQ....	14
NRG1_beta	....TGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCKCPNEFTGDRCQNYVMASFY....	15
Translated genomic locus	....TGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCK*	73, 128
NRG2_alpha	....SWSGHARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPNGFFGQRCLEKLPLRLY....	16
NRG2_beta	....SWSGHARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPVGYTGDRCQQFAMVNFY....	17
Translated genomic locus	....SWSGHARKCNETAKSYCVNGGVCYYIEGINQLS---CK*	74, 129
NRG3	....ERSEHFKPCRDKDLAYCLNDGECFVIETLTGSH-KHCRCKEGYQGVRCDQ-FLPKTD....	18
Translated genomic locus	....ERSEHFKPCRDKDLAYCLNDGECFVIETLTGSH-KHCR*	75, 130
NRG4	MPTDHEEPCGPSHKSFCLNGGLCYVIPTIPSPF---CRCVENYTGARCEEVFLPGSS....	19
Translated genomic locus	MPTDHEEPCGPSHKSFCLNGGLCYVIPTIPSPF---CR*	76, 131
EGF	....SVRNSDSECPLSHDGYCLHDGVCMYIEALDKYA---CNCVVGYIGERCQYRDLKWWE....	20
Translated genomic locus	....SVRNSDSECPLSHDGYCLHDGVCMYIEALDKYA---CK*	77, 132
TGF_alpha	....AVVSHFNDCPDSHTQFCFH-GTCRFLVQEDKPA---CVCHSGYVGARCEHADLLAVV....	21
Translated genomic locus	....AVVSHFNDCPDSHTQFCFH-GTCRFLVQEDKPA---CV*	78, 133
Betacellulin	....KRKGHFSRCPKQYKHYCIK-GRCRFVVAEQTPS---CVCDEGYIGARCERVDLFYLR....	22
Translated genomic locus	....KRKGHFSRCPKQYKHYCIK-GRCRFVVAEQTPS---CV*	79, 134

FIGURE 4

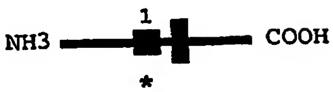
Amphiregulin	....RNRKKKNPCNAEFQNFCIH-GECKYIEHLEAVT---CKCQQEYFGERCGEKSMKTHS....	23
Translated genomic locus	....RNRKKKNPCNAEFQNFCIH-GECKYIEHLEAVT---CK*	80, 135
HB-EGF	....GLGKKRDPCLRKYKDFCIH-GECKYVKELRAPS---CTCHPGYHGERCHGLSLPVEN....	24
Translated genomic locus	....GLGKKRDPCLRKYKDFCIH-GECKYVKELRAPS---CM*	81, 136
Epiregulin	....VAQVSITKCSSDMNGYCLH-GQCIYLVDMSQNY---CRCEVGYTGVRCEHFFLTVHQ....	25
Translated genomic locus	....VAQVSITKCSSDMNGYCLH-GQCIYLVDMSQNY---CR*	82, 137
Epigen(Mouse)	....VALKFSHPCLEDHNSYCIN-GACAFHHELKQAI---CRCFTGYTGQRCEHLTLTSYA....	26
Trans. mouse genomic locus	....VALKFSHPCLEDHNSYCIN-GACAFHHELKQAI---CR*	83, 138
Trans. human genomic locus	....IALKFSHLCLEDHNSYCIN-GACAFHHELEKAI---CR*	84, 139

Loop A Loop B

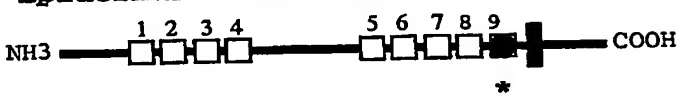
FIGURE 4 (continued)

Figure 5A

i) **TGF-alpha**



ii) **Epidermal Growth Factor**



iii) **Notch 1**

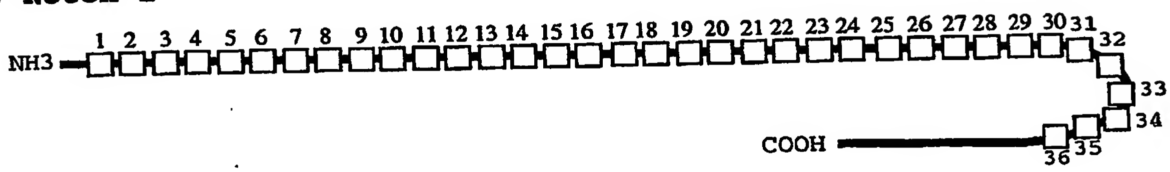


Figure 5B

i) **TGF alpha**

EGF DOMAIN NUMBER		----------------SEQUENCE----------------	Sequence ID #
1.	**EGF_47_82** *	CPDSHTQFCF-HGT-CRFLVQEDKPACVCHSG--YVGAR--C	72

ii) **EPIDERMAL GROWTH FACTOR**

EGF DOMAIN NUMBER		----------------SEQUENCE----------------	Sequence ID #
1.	EGF_318_354	CKLR-KGNCS--STVCGQDL--QSHLCLCAEGYALSRDRKYC	27
2.	EGF_360_395	CAFW-NHGC---TLGCKNTP--GSYYCTCPVGFVLLPDGKRC	28
3.	EGF_401_436	CPRN-VSEC---SHDCVLTS--EGPLCFCPEGSVLERDGKTC	29
4.	EGF_439-476	SSSPDNGGCS--QLCVPLSP--VSWECDCFPGYDLQLDEKSC	30
5.	EGF_745_780	CLYQ-NGGC---EHICKKRL--GTAVCSCREGFMKASDGKTC	31
6.	EGF_835_868	CAPV---CGS-MYARCISEG--EDATCQCLKG--FAGDGKLC	32
7.	EGF_874_910	CHMG-VPVCPPASSKCINTE--GGYVCRCSEG--YQGDGIHC	33
8.	EGF_916_951	CQLG-VHSCGENAS-CTNTE--GGYTCMCAGR--LSEPGLIC	34
9.	**EGF_976_1012** *	CPLSHDGYCL-HDGVCMYIEALDKYACNCVVG--YIGER--C	35

iii) NotchI

EGF DOMAIN NUMBER	SEQUENCE	Sequence ID #
1. EGF_24_57	CSQ---------PGETCLNGGKCEAANGTE------ACVCG-GAFVGPRC	36
2. EGF_63_98	CLS---------TPCKNAGTCH-VVDRRGVADYACSCA-LGFSGPLC	37
3. EGF_106_138	CLT---------NPCKNGGTCD-LLTLT-----EYKCRCP-PGWSGKSC	38
4. EGF_144_175	CAS---------NPCANGGQC----LP-FEASYICHCP-PSFHGPTC	39
5. EGF_182_215	CGQ---------KPRLCRHGGTCHNEVGSY-----RCVCR-ATHTGPNC	40
6. EGF_222_254	CSPSP---------CQNGGTCRPTGDV------THECACL-PGFTGQNC	41
7. EGF_261_292	CPDNN---------CKNGGACV-DGVN------TYNCRCP-PEWTGQYC	42
8. EGF_299_332	CQL-------MPNACQNGGTCHNTHGGY------NCVCV-NGWTGEDC	43
9. EGF_339_370	CAS---------AACFHGATCH-DRVA------SFYCECP-HGRTGLLC	44
10. EGF_376_409	CIS---------NPCNEGSNCD-TNPV----NGKAICTCP-SGYTGPAC	45
11. EGF_416_449	CSL---------GANPCEHAGKCINTLGSF------ECQCL-QGYTGPRC	46
12. EGF_456_487	CVS---------NPCQNDATCL-DQIG------EFQCICM-PGYEGVHC	47
13. EGF_494_525	CAS---------SPCLHNGRCL-DKIN------EFQCECP-TGFTGHLC	48
14. EGF_532_563	CAS---------TPCKNGAKCL-DGPN------TYTCVCT-EGYTGTHC	49
15. EGF_570_600	CDPDP---------CHYG-SCK-DGVA------TFTCLCR-PGYTGHHC	50
16. EGF_607_638	CSS---------QPCRLRGTCQ-DPDN------AYLCFCL-KGTTGPNC	51
17. EGF_645_675	CAS---------SPC-DSGTCL-DKID------GYECACE-PGYTGSMC	52
18. EGF_682_713	CAG---------NPCHNGGTC-----ED-GINGFTCRCP-EGYHDPTC	53
19. EGF_720_750	CNS---------NPCVHGACRD----SLN----GYKCDCD-PGWSGTNC	54
20. EGF_757_788	CES---------NPCVNGGTCK-D-MTS-----GIVCTCR-EGFSGPNC	55
21. EGF_795_826	CAS---------NPCLNKGTC-----IDD----VAGYKCNCL-LPYTGATC	56
22. EGF_833_867	CAP---------SPCRNGGECR-QSED-YES-FSCVCPT-AGAKGQTC	57
23. EGF_874_905	CVL---------SPCRHGASCQ-NTHG------GYRCHCQ-AGYSGRNC	58
24. EGF_912_943	CRPNP-------------CHNGGSCT-DGIN------TAFCDCL-PGFRGTFC	59
25. EGF_950_981	CAS--DP---------CRNGANCT-DCVD------SYTCTCP-AGFSGIHC	60
26. EGF_988_1019	CTESS-------------CFNGGTCV-DGIN------SFTCLCP-PGFTGSYC	61
27. EGF_1026_1057	CDS---------RPCLLGGTCQ-DG------RGLHRCTCP-QGYTGPNC	62
28. EGF_1064_1095	CDS---------SPCKNGGKCW-QT------HTQYRCECP-SGWTGLYC	63
29. EGF_1102_1143	CEVAAQRQGVDVARLCQHGGLCV-DAGN------THHCRCQ-AGYTGSYC	64
30. EGF_1150_1181	CSP---------SPCQNGATCT-DYLG------GYSCKCV-AGYHGVNC	65
31. EGF_1188_1219	CLS---------HPCQNGGTCL-DLPN------TYKCSCP-RGTQGVHC	66
32. EGF_1226_1265	CNPPVDPVSRS--PKCFNNGTCV-DQVG------GYSCTCP-PGFVGERC	67
33. EGF_1272_1305	CLS---------NPCDARGTQN-CVQR------VNDFHCECR-AGHTGRRC	68
34. EGF_1312_1346	CKG---------KPCKNGGTCA-VASN-TARGFICKCP-AGFEGATC	69
35. EGF_1353_1384	C---------GSLRCLNGGTCISGPRSP------TCLCL-GPFTGPEC	70
36. EGF_1392_1426	CL---------GGNPCYNQGTCEPTSESPF------YRCLCP-AKFNGLLC	71

SEQUENCE LISTING

<110> Harari, Daniel

<120> SPLICE VARIANTS OF ERB-B RECEPTOR LIGANDS, COMPOSITIONS AND USES THEREOF

<130> Harari-001

<160> 182

<170> PatentIn version 3.1

<210> 1
<211> 56
<212> PRT
<213> Homo sapiens

<400> 1

```
Thr Gly Thr Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe
1               5                   10                  15

Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro
            20                  25                  30

Ser Arg Tyr Leu Cys Lys Cys Gln Pro Gly Phe Thr Gly Ala Arg Cys
        35                  40                  45

Thr Glu Asn Val Pro Met Lys Val
    50                  55
```

<210> 2
<211> 56
<212> PRT
<213> Homo sapiens

<400> 2

```
Thr Gly Thr Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe
1               5                   10                  15

Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro
            20                  25                  30

Ser Arg Tyr Leu Cys Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys
        35                  40                  45

Gln Asn Tyr Val Met Ala Ser Phe
    50                  55
```

<210> 3
<211> 53
<212> PRT
<213> Homo sapiens

<400> 3

Ser Trp Ser Gly His Ala Arg Lys Cys Asn Glu Thr Ala Lys Ser Tyr
1   5   10   15

Cys Val Asn Gly Gly Val Cys Tyr Tyr Ile Glu Gly Ile Asn Gln Leu
            20   25   30

Ser Cys Lys Cys Pro Asn Gly Phe Phe Gly Gln Arg Cys Leu Glu Lys
        35   40   45

Leu Pro Leu Arg Leu
    50

<210> 4
<211> 53
<212> PRT
<213> Homo sapiens

<400> 4

Ser Trp Ser Gly His Ala Arg Lys Cys Asn Glu Thr Ala Lys Ser Tyr
1   5   10   15

Cys Val Asn Gly Gly Val Cys Tyr Tyr Ile Glu Gly Ile Asn Gln Leu
            20   25   30

Ser Cys Lys Cys Pro Val Gly Tyr Thr Gly Asp Arg Cys Gln Gln Phe
        35   40   45

Ala Met Val Asn Phe
    50

<210> 5
<211> 55
<212> PRT
<213> Homo sapiens

<400> 5

Glu Arg Ser Glu His Phe Lys Pro Cys Arg Asp Lys Asp Leu Ala Tyr
1   5   10   15

Cys Leu Asn Asp Gly Glu Cys Phe Val Ile Glu Thr Leu Thr Gly Ser
            20   25   30

His Lys His Cys Arg Cys Lys Glu Gly Tyr Gln Gly Val Arg Cys Asp
        35   40   45

Gln Phe Leu Pro Lys Thr Asp
    50   55

<210> 6
<211> 53
<212> PRT
<213> Homo sapiens

<400> 6

```
Met Pro Thr Asp His Glu Glu Pro Cys Gly Pro Ser His Lys Ser Phe
1               5                   10                  15

Cys Leu Asn Gly Gly Leu Cys Tyr Val Ile Pro Thr Ile Pro Ser Pro
            20                  25                  30

Phe Cys Arg Cys Val Glu Asn Tyr Thr Gly Ala Arg Cys Glu Glu Val
        35                  40                  45

Phe Leu Pro Gly Ser
    50
```

<210> 7
<211> 53
<212> PRT
<213> Homo sapiens

<400> 7

```
Ser Val Arg Asn Ser Asp Ser Glu Cys Pro Leu Ser His Asp Gly Tyr
1               5                   10                  15

Cys Leu His Asp Gly Val Cys Met Tyr Ile Glu Ala Leu Asp Lys Tyr
            20                  25                  30

Ala Cys Asn Cys Val Val Gly Tyr Ile Gly Glu Arg Cys Gln Tyr Arg
        35                  40                  45

Asp Leu Lys Trp Trp
    50
```

<210> 8
<211> 52
<212> PRT
<213> Homo sapiens

<400> 8

```
Ala Val Val Ser His Phe Asn Asp Cys Pro Asp Ser His Thr Gln Phe
1               5                   10                  15

Cys Phe His Gly Thr Cys Arg Phe Leu Val Gln Glu Asp Lys Pro Ala
            20                  25                  30

Cys Val Cys His Ser Gly Tyr Val Gly Ala Arg Cys Glu His Ala Asp
        35                  40                  45

Leu Leu Ala Val
    50
```

<210> 9

<211> 52
<212> PRT
<213> Homo sapiens

<400> 9

Lys Arg Lys Gly His Phe Ser Arg Cys Pro Lys Gln Tyr Lys His Tyr
1   5   10   15

Cys Ile Lys Gly Arg Cys Arg Phe Val Val Ala Glu Gln Thr Pro Ser
            20   25   30

Cys Val Cys Asp Glu Gly Tyr Ile Gly Ala Arg Cys Glu Arg Val Asp
        35   40   45

Leu Phe Tyr Leu
    50

<210> 10
<211> 52
<212> PRT
<213> Homo sapiens

<400> 10

Arg Asn Arg Lys Lys Lys Asn Pro Cys Asn Ala Glu Phe Gln Asn Phe
1   5   10   15

Cys Ile His Gly Glu Cys Lys Tyr Ile Glu His Leu Glu Ala Val Thr
            20   25   30

Cys Lys Cys Gln Gln Glu Tyr Phe Gly Glu Arg Cys Gly Glu Lys Ser
        35   40   45

Met Lys Thr His
    50

<210> 11
<211> 52
<212> PRT
<213> Homo sapiens

<400> 11

Gly Leu Gly Lys Lys Arg Asp Pro Cys Leu Arg Lys Tyr Lys Asp Phe
1   5   10   15

Cys Ile His Gly Glu Cys Lys Tyr Val Lys Glu Leu Arg Ala Pro Ser
            20   25   30

Cys Ile Cys His Pro Gly Tyr His Gly Glu Arg Cys His Gly Leu Ser
        35   40   45

Leu Pro Val Glu
    50

<210> 12
<211> 52
<212> PRT
<213> Homo sapiens

<400> 12

Val Ala Gln Val Ser Ile Thr Lys Cys Ser Ser Asp Met Asn Gly Tyr
1   5   10   15

Cys Leu His Gly Gln Cys Ile Tyr Leu Val Asp Met Ser Gln Asn Tyr
            20   25   30

Cys Arg Cys Glu Val Gly Tyr Thr Gly Val Arg Cys Glu His Phe Phe
        35   40   45

Leu Thr Val His
    50

<210> 13
<211> 52
<212> PRT
<213> Mus musculus

<400> 13

Val Ala Leu Lys Phe Ser His Pro Cys Leu Glu Asp His Asn Ser Tyr
1   5   10   15

Cys Ile Asn Gly Ala Cys Ala Phe His His Glu Leu Lys Gln Ala Ile
            20   25   30

Cys Arg Cys Phe Thr Gly Tyr Thr Gly Gln Arg Cys Glu His Leu Thr
        35   40   45

Leu Thr Ser Tyr
    50

<210> 14
<211> 57
<212> PRT
<213> Homo sapiens

<400> 14

Thr Gly Thr Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe
1   5   10   15

Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro
            20   25   30

Ser Arg Tyr Leu Cys Lys Cys Gln Pro Gly Phe Thr Gly Ala Arg Cys
        35   40   45

Thr Glu Asn Val Pro Met Lys Val Gln
50 55

<210> 15
<211> 57
<212> PRT
<213> Homo sapiens

<400> 15

Thr Gly Thr Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe
1 5 10 15

Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro
20 25 30

Ser Arg Tyr Leu Cys Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys
35 40 45

Gln Asn Tyr Val Met Ala Ser Phe Tyr
50 55

<210> 16
<211> 54
<212> PRT
<213> Homo sapiens

<400> 16

Ser Trp Ser Gly His Ala Arg Lys Cys Asn Glu Thr Ala Lys Ser Tyr
1 5 10 15

Cys Val Asn Gly Gly Val Cys Tyr Tyr Ile Glu Gly Ile Asn Gln Leu
20 25 30

Ser Cys Lys Cys Pro Asn Gly Phe Phe Gly Gln Arg Cys Leu Glu Lys
35 40 45

Leu Pro Leu Arg Leu Tyr
50

<210> 17
<211> 54
<212> PRT
<213> Homo sapiens

<400> 17

Ser Trp Ser Gly His Ala Arg Lys Cys Asn Glu Thr Ala Lys Ser Tyr
1 5 10 15

Cys Val Asn Gly Gly Val Cys Tyr Tyr Ile Glu Gly Ile Asn Gln Leu
20 25 30

Ser Cys Lys Cys Pro Val Gly Tyr Thr Gly Asp Arg Cys Gln Gln Phe
35 40 45

Ala Met Val Asn Phe Tyr
50

<210> 18
<211> 55
<212> PRT
<213> Homo sapiens

<400> 18

Glu Arg Ser Glu His Phe Lys Pro Cys Arg Asp Lys Asp Leu Ala Tyr
1 5 10 15

Cys Leu Asn Asp Gly Glu Cys Phe Val Ile Glu Thr Leu Thr Gly Ser
20 25 30

His Lys His Cys Arg Cys Lys Glu Gly Tyr Gln Gly Val Arg Cys Asp
35 40 45

Gln Phe Leu Pro Lys Thr Asp
50 55

<210> 19
<211> 54
<212> PRT
<213> Homo sapiens

<400> 19

Met Pro Thr Asp His Glu Glu Pro Cys Gly Pro Ser His Lys Ser Phe
1 5 10 15

Cys Leu Asn Gly Gly Leu Cys Tyr Val Ile Pro Thr Ile Pro Ser Pro
20 25 30

Phe Cys Arg Cys Val Glu Asn Tyr Thr Gly Ala Arg Cys Glu Glu Val
35 40 45

Phe Leu Pro Gly Ser Ser
50

<210> 20
<211> 54
<212> PRT
<213> Homo sapiens

<400> 20

Ser Val Arg Asn Ser Asp Ser Glu Cys Pro Leu Ser His Asp Gly Tyr
1 5 10 15

Cys Leu His Asp Gly Val Cys Met Tyr Ile Glu Ala Leu Asp Lys Tyr

20 25 30

Ala Cys Asn Cys Val Val Gly Tyr Ile Gly Glu Arg Cys Gln Tyr Arg
35 40 45

Asp Leu Lys Trp Trp Glu
50

<210> 21
<211> 53
<212> PRT
<213> Homo sapiens

<400> 21

Ala Val Val Ser His Phe Asn Asp Cys Pro Asp Ser His Thr Gln Phe
1 5 10 15

Cys Phe His Gly Thr Cys Arg Phe Leu Val Gln Glu Asp Lys Pro Ala
20 25 30

Cys Val Cys His Ser Gly Tyr Val Gly Ala Arg Cys Glu His Ala Asp
35 40 45

Leu Leu Ala Val Val
50

<210> 22
<211> 53
<212> PRT
<213> Homo sapiens

<400> 22

Lys Arg Lys Gly His Phe Ser Arg Cys Pro Lys Gln Tyr Lys His Tyr
1 5 10 15

Cys Ile Lys Gly Arg Cys Arg Phe Val Val Ala Glu Gln Thr Pro Ser
20 25 30

Cys Val Cys Asp Glu Gly Tyr Ile Gly Ala Arg Cys Glu Arg Val Asp
35 40 45

Leu Phe Tyr Leu Arg
50

<210> 23
<211> 53
<212> PRT
<213> Homo sapiens

<400> 23

Arg Asn Arg Lys Lys Lys Asn Pro Cys Asn Ala Glu Phe Gln Asn Phe
1 5 10 15

Cys Ile His Gly Glu Cys Lys Tyr Ile Glu His Leu Glu Ala Val Thr
20 25 30

Cys Lys Cys Gln Gln Glu Tyr Phe Gly Glu Arg Cys Gly Glu Lys Ser
35 40 45

Met Lys Thr His Ser
50

<210> 24
<211> 53
<212> PRT
<213> Homo sapiens

<400> 24

Gly Leu Gly Lys Lys Arg Asp Pro Cys Leu Arg Lys Tyr Lys Asp Phe
1 5 10 15

Cys Ile His Gly Glu Cys Lys Tyr Val Lys Glu Leu Arg Ala Pro Ser
20 25 30

Cys Ile Cys His Pro Gly Tyr His Gly Glu Arg Cys His Gly Leu Ser
35 40 45

Leu Pro Val Glu Asn
50

<210> 25
<211> 53
<212> PRT
<213> Homo sapiens

<400> 25

Val Ala Gln Val Ser Ile Thr Lys Cys Ser Ser Asp Met Asn Gly Tyr
1 5 10 15

Cys Leu His Gly Gln Cys Ile Tyr Leu Val Asp Met Ser Gln Asn Tyr
20 25 30

Cys Arg Cys Glu Val Gly Tyr Thr Gly Val Arg Cys Glu His Phe Phe
35 40 45

Leu Thr Val His Gln
50

<210> 26
<211> 53
<212> PRT
<213> Mus musculus

<400> 26

Val Ala Leu Lys Phe Ser His Pro Cys Leu Glu Asp His Asn Ser Tyr
1 5 10 15

Cys Ile Asn Gly Ala Cys Ala Phe His His Glu Leu Lys Gln Ala Ile
20 25 30

Cys Arg Cys Phe Thr Gly Tyr Thr Gly Gln Arg Cys Glu His Leu Thr
35 40 45

Leu Thr Ser Tyr Ala
50

<210> 27
<211> 37
<212> PRT
<213> Homo sapiens

<400> 27

Cys Lys Leu Arg Lys Gly Asn Cys Ser Ser Thr Val Cys Gly Gln Asp
1 5 10 15

Leu Gln Ser His Leu Cys Met Cys Ala Glu Gly Tyr Ala Leu Ser Arg
20 25 30

Asp Arg Lys Tyr Cys
35

<210> 28
<211> 36
<212> PRT
<213> Homo sapiens

<400> 28

Cys Ala Phe Trp Asn His Gly Cys Thr Leu Gly Cys Lys Asn Thr Pro
1 5 10 15

Gly Ser Tyr Tyr Cys Thr Cys Pro Val Gly Phe Val Leu Leu Pro Asp
20 25 30

Gly Lys Arg Cys
35

<210> 29
<211> 36
<212> PRT
<213> Homo sapiens

<400> 29

Cys Pro Arg Asn Val Ser Glu Cys Ser His Asp Cys Val Leu Thr Ser
1 5 10 15

Glu Gly Pro Leu Cys Phe Cys Pro Glu Gly Ser Val Leu Glu Arg Asp
            20   25   30

Gly Lys Thr Cys
        35

<210> 30
<211> 38
<212> PRT
<213> Homo sapiens

<400> 30

Cys Ser Ser Pro Asp Asn Gly Gly Cys Ser Gln Leu Cys Val Pro Leu
1   5   10   15

Ser Pro Val Ser Trp Glu Cys Asp Cys Phe Pro Gly Tyr Asp Leu Gln
            20   25   30

Leu Asp Glu Lys Ser Cys
        35

<210> 31
<211> 36
<212> PRT
<213> Homo sapiens

<400> 31

Cys Leu Tyr Gln Asn Gly Gly Cys Glu His Ile Cys Lys Lys Arg Leu
1   5   10   15

Gly Thr Ala Trp Cys Ser Cys Arg Glu Gly Phe Met Lys Ala Ser Asp
            20   25   30

Gly Lys Thr Cys
        35

<210> 32
<211> 34
<212> PRT
<213> Homo sapiens

<400> 32

Cys Ala Pro Val Gly Cys Ser Met Tyr Ala Arg Cys Ile Ser Glu Gly
1   5   10   15

Glu Asp Ala Thr Cys Gln Cys Leu Lys Gly Phe Ala Gly Asp Gly Lys
            20   25   30

Leu Cys

<210> 33

<211> 37
<212> PRT
<213> Homo sapiens

<400> 33

Cys Glu Met Gly Val Pro Val Cys Pro Pro Ala Ser Ser Lys Cys Ile
1 5 10 15

Asn Thr Glu Gly Gly Tyr Val Cys Arg Cys Ser Glu Gly Tyr Gln Gly
20 25 30

Asp Gly Ile His Cys
35

<210> 34
<211> 36
<212> PRT
<213> Homo sapiens

<400> 34

Cys Gln Leu Gly Val His Ser Cys Gly Glu Asn Ala Ser Cys Thr Asn
1 5 10 15

Thr Glu Gly Gly Tyr Thr Cys Met Cys Ala Gly Arg Leu Ser Glu Pro
20 25 30

Gly Leu Ile Cys
35

<210> 35
<211> 37
<212> PRT
<213> Homo sapiens

<400> 35

Cys Pro Leu Ser His Asp Gly Tyr Cys Leu His Asp Gly Val Cys Met
1 5 10 15

Tyr Ile Glu Ala Leu Asp Lys Tyr Ala Cys Asn Cys Val Val Gly Tyr
20 25 30

Ile Gly Glu Arg Cys
35

<210> 36
<211> 34
<212> PRT
<213> Homo sapiens

<400> 36

Cys Ser Gln Pro Gly Glu Thr Cys Leu Asn Gly Gly Lys Cys Glu Ala
1 5 10 15

Ala Asn Gly Thr Glu Ala Cys Val Cys Gly Gly Ala Phe Val Gly Pro
20 25 30

Arg Cys

<210> 37
<211> 36
<212> PRT
<213> Homo sapiens

<400> 37

Cys Leu Ser Thr Pro Cys Lys Asn Ala Gly Thr Cys His Val Val Asp
1 5 10 15

Arg Arg Gly Val Ala Asp Tyr Ala Cys Ser Cys Ala Leu Gly Phe Ser
20 25 30

Gly Pro Leu Cys
35

<210> 38
<211> 33
<212> PRT
<213> Homo sapiens

<400> 38

Cys Leu Thr Asn Pro Cys Arg Asn Gly Gly Thr Cys Asp Leu Leu Thr
1 5 10 15

Leu Thr Glu Tyr Lys Cys Arg Cys Pro Pro Gly Trp Ser Gly Lys Ser
20 25 30

Cys

<210> 39
<211> 32
<212> PRT
<213> Homo sapiens

<400> 39

Cys Ala Ser Asn Pro Cys Ala Asn Gly Gly Gln Cys Leu Pro Phe Glu
1 5 10 15

Ala Ser Tyr Ile Cys His Cys Pro Pro Ser Phe His Gly Pro Thr Cys
20 25 30

<210> 40
<211> 34
<212> PRT
<213> Homo sapiens

<400> 40

Cys Gly Gln Lys Pro Arg Leu Cys Arg His Gly Gly Thr Cys His Asn
1   5   10   15

Glu Val Gly Ser Tyr Arg Cys Val Cys Arg Ala Thr His Thr Gly Pro
            20   25   30

Asn Cys

<210> 41
<211> 33
<212> PRT
<213> Homo sapiens

<400> 41

Cys Ser Pro Ser Pro Cys Gln Asn Gly Gly Thr Cys Arg Pro Thr Gly
1   5   10   15

Asp Val Thr His Glu Cys Ala Cys Leu Pro Gly Phe Thr Gly Gln Asn
            20   25   30

Cys

<210> 42
<211> 32
<212> PRT
<213> Homo sapiens

<400> 42

Cys Pro Gly Asn Asn Cys Lys Asn Gly Gly Ala Cys Val Asp Gly Val
1   5   10   15

Asn Thr Tyr Asn Cys Pro Cys Pro Pro Glu Trp Thr Gly Gln Tyr Cys
            20   25   30

<210> 43
<211> 34
<212> PRT
<213> Homo sapiens

<400> 43

Cys Gln Leu Met Pro Asn Ala Cys Gln Asn Gly Gly Thr Cys His Asn
1   5   10   15

Thr His Gly Gly Tyr Asn Cys Val Cys Val Asn Gly Trp Thr Gly Glu
            20   25   30

Asp Cys

<210> 44
<211> 32
<212> PRT
<213> Homo sapiens

<400> 44

Cys Ala Ser Ala Ala Cys Phe His Gly Ala Thr Cys His Asp Arg Val
1 5 10 15

Ala Ser Phe Tyr Cys Glu Cys Pro His Gly Arg Thr Gly Leu Leu Cys
20 25 30

<210> 45
<211> 34
<212> PRT
<213> Homo sapiens

<400> 45

Cys Ile Ser Asn Pro Cys Asn Glu Gly Ser Asn Cys Asp Thr Asn Pro
1 5 10 15

Val Asn Gly Lys Ala Ile Cys Thr Cys Pro Ser Gly Tyr Thr Gly Pro
20 25 30

Ala Cys

<210> 46
<211> 34
<212> PRT
<213> Homo sapiens

<400> 46

Cys Ser Leu Gly Ala Asn Pro Cys Glu His Ala Gly Lys Cys Ile Asn
1 5 10 15

Thr Leu Gly Ser Phe Glu Cys Gln Cys Leu Gln Gly Tyr Thr Gly Pro
20 25 30

Arg Cys

<210> 47
<211> 32
<212> PRT
<213> Mus musculus

<400> 47

Cys Val Ser Asn Pro Cys Gln Asn Asp Ala Thr Cys Leu Asp Gln Ile
1 5 10 15

Gly Glu Phe Gln Cys Met Cys Met Pro Gly Tyr Glu Gly Val His Cys
20 25 30

<210> 48
<211> 32
<212> PRT
<213> Homo sapiens

<400> 48

Cys Ala Ser Ser Pro Cys Leu His Asn Gly Arg Cys Leu Asp Lys Ile
1 5 10 15

Asn Glu Phe Gln Cys Glu Cys Pro Thr Gly Phe Thr Gly His Leu Cys
20 25 30

<210> 49
<211> 32
<212> PRT
<213> Homo sapiens

<400> 49

Cys Ala Ser Thr Pro Cys Lys Asn Gly Ala Lys Cys Leu Asp Gly Pro
1 5 10 15

Asn Thr Tyr Thr Cys Val Cys Thr Glu Gly Tyr Thr Gly Thr His Cys
20 25 30

<210> 50
<211> 31
<212> PRT
<213> Homo sapiens

<400> 50

Cys Asp Pro Asp Pro Cys His Tyr Gly Ser Cys Lys Asp Gly Val Ala
1 5 10 15

Thr Phe Thr Cys Leu Cys Arg Pro Gly Tyr Thr Gly His His Cys
20 25 30

<210> 51
<211> 32
<212> PRT
<213> Homo sapiens

<400> 51

Cys Ser Ser Gln Pro Cys Arg Leu Arg Gly Thr Cys Gln Asp Pro Asp
1 5 10 15

Asn Ala Tyr Leu Cys Phe Cys Leu Lys Gly Thr Thr Gly Pro Asn Cys
20 25 30

<210> 52
<211> 31

<212> PRT
<213> Homo sapiens

<400> 52

Cys Ala Ser Ser Pro Cys Asp Ser Gly Thr Cys Leu Asp Lys Ile Asp
1   5   10   15

Gly Tyr Glu Cys Ala Cys Glu Pro Gly Tyr Thr Gly Ser Met Cys
            20   25   30

<210> 53
<211> 32
<212> PRT
<213> Homo sapiens

<400> 53

Cys Ala Gly Asn Pro Cys His Asn Gly Gly Thr Cys Glu Asp Gly Ile
1   5   10   15

Asn Gly Phe Thr Cys Arg Cys Pro Glu Gly Tyr His Asp Pro Thr Cys
            20   25   30

<210> 54
<211> 31
<212> PRT
<213> Homo sapiens

<400> 54

Cys Asn Ser Asn Pro Cys Val His Gly Ala Cys Arg Asp Ser Leu Asn
1   5   10   15

Gly Tyr Lys Cys Asp Cys Asp Pro Gly Trp Ser Gly Thr Asn Cys
            20   25   30

<210> 55
<211> 32
<212> PRT
<213> Homo sapiens

<400> 55

Cys Glu Ser Asn Pro Cys Val Asn Gly Gly Thr Cys Lys Asp Met Thr
1   5   10   15

Ser Gly Ile Val Cys Thr Cys Arg Glu Gly Phe Ser Gly Pro Asn Cys
            20   25   30

<210> 56
<211> 32
<212> PRT
<213> Homo sapiens

<400> 56

Cys Ala Ser Asn Pro Cys Leu Asn Lys Gly Thr Cys Ile Asp Asp Val

1 5 10 15

Ala Gly Tyr Lys Cys Asn Cys Leu Leu Pro Tyr Thr Gly Ala Thr Cys
20 25 30

<210> 57
<211> 35
<212> PRT
<213> Homo sapiens

<400> 57

Cys Ala Pro Ser Pro Cys Arg Asn Gly Gly Glu Cys Arg Gln Ser Glu
1 5 10 15

Asp Tyr Glu Ser Phe Ser Cys Val Cys Pro Thr Ala Gly Ala Lys Gly
20 25 30

Gln Thr Cys
35

<210> 58
<211> 32
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (18)..(18)
<223> X = undefined amino acid

<400> 58

Cys Val Leu Ser Pro Cys Arg His Gly Ala Ser Cys Gln Asn Thr His
1 5 10 15

Gly Xaa Tyr Arg Cys His Cys Gln Ala Gly Tyr Ser Gly Arg Asn Cys
20 25 30

<210> 59
<211> 32
<212> PRT
<213> Homo sapiens

<400> 59

Cys Arg Pro Asn Pro Cys His Asn Gly Gly Ser Cys Thr Asp Gly Ile
1 5 10 15

Asn Thr Ala Phe Cys Asp Cys Leu Pro Gly Phe Arg Gly Thr Phe Cys
20 25 30

<210> 60
<211> 32
<212> PRT
<213> Homo sapiens

<400> 60

Cys Ala Ser Asp Pro Cys Arg Asn Gly Ala Asn Cys Thr Asp Cys Val
1 5 10 15

Asp Ser Tyr Thr Cys Thr Cys Pro Ala Gly Phe Ser Gly Ile His Cys
20 25 30

<210> 61
<211> 32
<212> PRT
<213> Homo sapiens

<400> 61

Cys Thr Glu Ser Ser Cys Phe Asn Gly Gly Thr Cys Val Asp Gly Ile
1 5 10 15

Asn Ser Phe Thr Cys Leu Cys Pro Pro Gly Phe Thr Gly Ser Tyr Cys
20 25 30

<210> 62
<211> 32
<212> PRT
<213> Homo sapiens

<400> 62

Cys Asp Ser Arg Pro Cys Leu Leu Gly Gly Thr Cys Gln Asp Gly Arg
1 5 10 15

Gly Leu His Arg Cys Thr Cys Pro Gln Gly Tyr Thr Gly Pro Asn Cys
20 25 30

<210> 63
<211> 32
<212> PRT
<213> Homo sapiens

<400> 63

Cys Asp Ser Ser Pro Cys Lys Asn Gly Gly Lys Cys Trp Gln Thr His
1 5 10 15

Thr Gln Tyr Arg Cys Glu Cys Pro Ser Gly Trp Thr Gly Leu Tyr Cys
20 25 30

<210> 64
<211> 42
<212> PRT
<213> Homo sapiens

<400> 64

Cys Glu Val Ala Ala Gln Arg Gln Gly Val Asp Val Ala Arg Leu Cys
1 5 10 15

Gln His Gly Gly Leu Cys Val Asp Ala Gly Asn Thr His His Cys Arg
20 25 30

Cys Gln Ala Gly Tyr Thr Gly Ser Tyr Cys
35 40

<210> 65
<211> 32
<212> PRT
<213> Homo sapiens

<400> 65

Cys Ser Pro Ser Pro Cys Gln Asn Gly Ala Thr Cys Thr Asp Tyr Leu
1 5 10 15

Gly Gly Tyr Ser Cys Lys Cys Val Ala Gly Tyr His Gly Val Asn Cys
20 25 30

<210> 66
<211> 32
<212> PRT
<213> Homo sapiens

<400> 66

Cys Leu Ser His Pro Cys Gln Asn Gly Gly Thr Cys Leu Asp Leu Pro
1 5 10 15

Asn Thr Tyr Lys Cys Ser Cys Pro Arg Gly Thr Gln Gly Val His Cys
20 25 30

<210> 67
<211> 40
<212> PRT
<213> Mus musculus

<400> 67

Cys Asn Pro Pro Val Asp Pro Val Ser Arg Ser Pro Lys Cys Phe Asn
1 5 10 15

Asn Gly Thr Cys Val Asp Gln Val Gly Gly Tyr Ser Cys Thr Cys Pro
20 25 30

Pro Gly Phe Val Gly Glu Arg Cys
35 40

<210> 68
<211> 34
<212> PRT
<213> Homo sapiens

<400> 68

Cys Leu Ser Asn Pro Cys Asp Ala Arg Gly Thr Gln Asn Cys Val Gln

1   5   10   15

Arg Val Asn Asp Phe His Cys Glu Cys Arg Ala Gly His Thr Gly Arg
            20   25   30

Arg Cys

<210> 69
<211> 35
<212> PRT
<213> Homo sapiens

<400> 69

Cys Lys Gly Lys Pro Cys Lys Asn Gly Gly Thr Cys Ala Val Ala Ser
1   5   10   15

Asn Thr Ala Arg Gly Phe Ile Cys Lys Cys Pro Ala Gly Phe Glu Gly
            20   25   30

Ala Thr Cys
        35

<210> 70
<211> 32
<212> PRT
<213> Homo sapiens

<400> 70

Cys Gly Ser Leu Arg Cys Leu Asn Gly Gly Thr Cys Ile Ser Gly Pro
1   5   10   15

Arg Ser Pro Thr Cys Leu Cys Leu Gly Pro Phe Thr Gly Pro Glu Cys
            20   25   30

<210> 71
<211> 35
<212> PRT
<213> Homo sapiens

<400> 71

Cys Leu Gly Gly Asn Pro Cys Tyr Asn Gln Gly Thr Cys Glu Pro Thr
1   5   10   15

Ser Glu Ser Pro Phe Tyr Arg Cys Leu Cys Pro Ala Lys Phe Asn Gly
            20   25   30

Leu Leu Cys
        35

<210> 72
<211> 36

<212> PRT
<213> Homo sapiens

<400> 72

Cys Pro Asp Ser His Thr Gln Phe Cys Phe His Gly Thr Cys Arg Phe
1   5   10   15

Leu Val Gln Glu Asp Lys Pro Ala Cys Val Cys His Ser Gly Tyr Val
            20   25   30

Gly Ala Arg Cys
        35

<210> 73
<211> 38
<212> PRT
<213> Homo sapiens

<400> 73

Thr Gly Thr Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe
1   5   10   15

Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro
            20   25   30

Ser Arg Tyr Leu Cys Lys
        35

<210> 74
<211> 35
<212> PRT
<213> Homo sapiens

<400> 74

Ser Trp Ser Gly His Ala Arg Lys Cys Asn Glu Thr Ala Lys Ser Tyr
1   5   10   15

Cys Val Asn Gly Gly Val Cys Tyr Tyr Ile Glu Gly Ile Asn Gln Leu
            20   25   30

Ser Cys Lys
        35

<210> 75
<211> 37
<212> PRT
<213> Homo sapiens

<400> 75

Glu Arg Ser Glu His Phe Lys Pro Cys Arg Asp Lys Asp Leu Ala Tyr
1   5   10   15

Cys Leu Asn Asp Gly Glu Cys Phe Val Ile Glu Thr Leu Thr Gly Ser
20 25 30

His Lys His Cys Arg
35

<210> 76
<211> 35
<212> PRT
<213> Homo sapiens

<400> 76

Met Pro Thr Asp His Glu Glu Pro Cys Gly Pro Ser His Lys Ser Phe
1 5 10 15

Cys Leu Asn Gly Gly Leu Cys Tyr Val Ile Pro Thr Ile Pro Ser Pro
20 25 30

Phe Cys Arg
35

<210> 77
<211> 35
<212> PRT
<213> Homo sapiens

<400> 77

Ser Val Arg Asn Ser Asp Ser Glu Cys Pro Leu Ser His Asp Gly Tyr
1 5 10 15

Cys Leu His Asp Gly Val Cys Met Tyr Ile Glu Ala Leu Asp Lys Tyr
20 25 30

Ala Cys Lys
35

<210> 78
<211> 34
<212> PRT
<213> Homo sapiens

<400> 78

Ala Val Val Ser His Phe Asn Asp Cys Pro Asp Ser His Thr Gln Phe
1 5 10 15

Cys Phe His Gly Thr Cys Arg Phe Leu Val Gln Glu Asp Lys Pro Ala
20 25 30

Cys Val

<210> 79

<211> 34
<212> PRT
<213> Homo sapiens

<400> 79

Lys Arg Lys Gly His Phe Ser Arg Cys Pro Lys Gln Tyr Lys His Tyr
1 5 10 15

Cys Ile Lys Gly Arg Cys Arg Phe Val Val Ala Glu Gln Thr Pro Ser
20 25 30

Cys Val

<210> 80
<211> 34
<212> PRT
<213> Homo sapiens

<400> 80

Arg Asn Arg Lys Lys Lys Asn Pro Cys Asn Ala Glu Phe Gln Asn Phe
1 5 10 15

Cys Ile His Gly Glu Cys Lys Tyr Ile Glu His Leu Glu Ala Val Thr
20 25 30

Cys Lys

<210> 81
<211> 34
<212> PRT
<213> Homo sapiens

<400> 81

Gly Leu Gly Lys Lys Arg Asp Pro Cys Leu Arg Lys Tyr Lys Asp Phe
1 5 10 15

Cys Ile His Gly Glu Cys Lys Tyr Val Lys Glu Leu Arg Ala Pro Ser
20 25 30

Cys Met

<210> 82
<211> 34
<212> PRT
<213> Homo sapiens

<400> 82

Val Ala Gln Val Ser Ile Thr Lys Cys Ser Ser Asp Met Asn Gly Tyr
1 5 10 15

Cys Leu His Gly Gln Cys Ile Tyr Leu Val Asp Met Ser Gln Asn Tyr
20 25 30

Cys Arg

<210> 83
<211> 34
<212> PRT
<213> Mus musculus

<400> 83

Val Ala Leu Lys Phe Ser His Pro Cys Leu Glu Asp His Asn Ser Tyr
1 5 10 15

Cys Ile Asn Gly Ala Cys Ala Phe His His Glu Leu Lys Gln Ala Ile
20 25 30

Cys Arg

<210> 84
<211> 34
<212> PRT
<213> Homo sapiens

<400> 84

Ile Ala Leu Lys Phe Ser His Leu Cys Leu Glu Asp His Asn Ser Tyr
1 5 10 15

Cys Ile Asn Gly Ala Cys Ala Phe His His Glu Leu Glu Lys Ala Ile
20 25 30

Cys Arg

<210> 85
<211> 360
<212> PRT
<213> Homo sapiens

<400> 85

Thr Ala Arg Gly Ala Gly Glu Glu Phe Pro Glu Thr Cys Trp Asn Ser
1 5 10 15

Gly Leu Ala Arg Arg Pro Gly Ala Glu Arg Arg Arg Leu Pro Asp Asp
20 25 30

Gly Ser Val Ser Arg Thr Val Ile Thr Ser Pro Arg Ser Gly Cys Glu
35 40 45

Gly Ala Gly Gln Arg Pro Gly Arg Glu Pro Pro Ala Ala Gly Pro Ile
50 55 60

Asp Asp Phe Pro Gly Arg Gln Glu Gln Pro Arg Glu Pro Gly Arg Ala
65 70 75 80

Pro Val Pro Gly Gly Arg Thr Ala Arg Arg Val Arg Ala Ala Leu Pro
85 90 95

Ala Gly Asn Gly Arg Arg Pro Arg Ala Ala Arg Ala Pro Gln Arg Gly
100 105 110

Arg Ser Leu Ser Pro Ser Arg Asp Lys Leu Phe Pro Asn Pro Ile Arg
115 120 125

Ala Leu Gly Pro Asn Ser Pro Ala Pro Arg Ala Val Arg Val Glu Arg
130 135 140

Ser Val Ser Gly Glu Met Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly
145 150 155 160

Lys Gly Lys Lys Lys Glu Arg Gly Ser Gly Lys Lys Pro Glu Ser Ala
165 170 175

Ala Gly Ser Gln Ser Pro Ala Leu Pro Pro Gln Leu Lys Glu Met Lys
180 185 190

Ser Gln Glu Ser Ala Ala Gly Ser Lys Leu Val Leu Arg Cys Glu Thr
195 200 205

Ser Ser Glu Tyr Ser Ser Leu Arg Phe Lys Trp Phe Lys Asn Gly Asn
210 215 220

Glu Leu Asn Arg Lys Asn Lys Pro Gln Asn Ile Lys Ile Gln Lys Lys
225 230 235 240

Pro Gly Lys Ser Glu Leu Arg Ile Asn Lys Ala Ser Leu Ala Asp Ser
245 250 255

Gly Glu Tyr Met Cys Lys Val Ile Ser Lys Leu Gly Asn Asp Ser Ala
260 265 270

Ser Ala Asn Ile Thr Ile Val Glu Ser Asn Glu Ile Ile Thr Gly Met
275 280 285

Pro Ala Ser Thr Glu Gly Ala Tyr Val Ser Ser Glu Ser Pro Ile Arg
290 295 300

Ile Ser Val Ser Thr Glu Gly Ala Asn Thr Ser Ser Ser Thr Ser Thr
305 310 315 320

Ser Thr Thr Gly Thr Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys
325 330 335

Thr Phe Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser
340 345 350

Asn Pro Ser Arg Tyr Leu Cys Lys
355 360

<210> 86
<211> 43
<212> PRT
<213> Homo sapiens

<400> 86

Thr Ser Thr Ser Thr Thr Gly Thr Ser His Leu Val Lys Cys Ala Glu
1 5 10 15

Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys
20 25 30

Asp Leu Ser Asn Pro Ser Arg Tyr Leu Cys Lys
35 40

<210> 87
<211> 43
<212> PRT
<213> Homo sapiens

<400> 87

Thr Ser Thr Ser Thr Thr Gly Thr Ser His Leu Val Lys Cys Ala Glu
1 5 10 15

Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys
20 25 30

Asp Leu Ser Asn Pro Ser Arg Tyr Leu Cys Lys
35 40

<210> 88
<211> 211
<212> PRT
<213> Homo sapiens

<400> 88

Met Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys Lys Lys
1 5 10 15

Glu Arg Gly Ser Gly Lys Lys Pro Glu Ser Ala Ala Gly Ser Gln Ser
20 25 30

Pro Ala Leu Pro Pro Gln Leu Lys Glu Met Lys Ser Gln Glu Ser Ala
35 40 45

Ala Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser Glu Tyr Ser
50 55 60

Ser Leu Arg Phe Lys Trp Phe Lys Asn Gly Asn Glu Leu Asn Arg Lys
65 70 75 80

Asn Lys Pro Gln Asn Ile Lys Ile Gln Lys Lys Pro Gly Lys Ser Glu
85 90 95

Leu Arg Ile Asn Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met Cys
100 105 110

Lys Val Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser Ala Asn Ile Thr
115 120 125

Ile Val Glu Ser Asn Glu Ile Ile Thr Gly Met Pro Ala Ser Thr Glu
130 135 140

Gly Ala Tyr Val Ser Ser Glu Ser Pro Ile Arg Ile Ser Val Ser Thr
145 150 155 160

Glu Gly Ala Asn Thr Ser Ser Ser Thr Ser Thr Ser Thr Thr Gly Thr
165 170 175

Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
180 185 190

Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
195 200 205

Leu Cys Lys
210

<210> 89
<211> 211
<212> PRT
<213> Homo sapiens

<400> 89

Met Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys Lys Lys
1 5 10 15

Glu Arg Gly Ser Gly Lys Lys Pro Glu Ser Ala Ala Gly Ser Gln Ser
20 25 30

Pro Ala Leu Pro Pro Gln Leu Lys Glu Met Lys Ser Gln Glu Ser Ala
35 40 45

Ala Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser Glu Tyr Ser
50 55 60

Ser Leu Arg Phe Lys Trp Phe Lys Asn Gly Asn Glu Leu Asn Arg Lys
65 70 75 80

Asn Lys Pro Gln Asn Ile Lys Ile Gln Lys Lys Pro Gly Lys Ser Glu
85 90 95

Leu Arg Ile Asn Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met Cys
100 105 110

Lys Val Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser Ala Asn Ile Thr
115 120 125

Ile Val Glu Ser Asn Glu Ile Ile Thr Gly Met Pro Ala Ser Thr Glu
130 135 140

Gly Ala Tyr Val Ser Ser Glu Ser Pro Ile Arg Ile Ser Val Ser Thr
145 150 155 160

Glu Gly Ala Asn Thr Ser Ser Ser Thr Ser Thr Ser Thr Thr Gly Thr
165 170 175

Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
180 185 190

Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
195 200 205

Leu Cys Lys
210

<210> 90
<211> 211
<212> PRT
<213> Mus musculus

<400> 90

Met Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys Lys Lys
1 5 10 15

Asp Arg Gly Ser Arg Gly Lys Pro Ala Pro Ala Glu Gly Asp Pro Ser
20 25 30

Pro Ala Leu Pro Pro Arg Leu Lys Glu Met Lys Ser Gln Glu Ser Ala
35 40 45

Ala Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser Glu Tyr Ser

50 55 60

Ser Leu Arg Phe Lys Trp Phe Lys Asn Gly Asn Glu Leu Asn Arg Arg
65 70 75 80

Asn Lys Pro Gln Asn Val Lys Ile Gln Lys Lys Pro Gly Lys Ser Glu
85 90 95

Leu Arg Ile Asn Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met Cys
100 105 110

Lys Val Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser Ala Asn Ile Thr
115 120 125

Ile Val Glu Ser Asn Asp Leu Thr Thr Gly Met Ser Ala Ser Thr Glu
130 135 140

Arg Pro Tyr Val Ser Ser Glu Ser Pro Ile Arg Ile Ser Val Ser Thr
145 150 155 160

Glu Gly Ala Asn Thr Ser Ser Ser Thr Ser Thr Ser Thr Thr Gly Thr
165 170 175

Ser His Leu Ile Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
180 185 190

Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
195 200 205

Leu Cys Lys
210

<210> 91
<211> 211
<212> PRT
<213> Mus musculus

<400> 91

Met Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys Lys Lys
1 5 10 15

Asp Arg Gly Ser Arg Gly Lys Pro Ala Pro Ala Glu Gly Asp Pro Ser
20 25 30

Pro Ala Leu Pro Pro Arg Leu Lys Glu Met Lys Ser Gln Glu Ser Ala
35 40 45

Ala Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser Glu Tyr Ser
50 55 60

Ser Leu Arg Phe Lys Trp Phe Lys Asn Gly Asn Glu Leu Asn Arg Arg
65   70   75   80

Asn Lys Pro Gln Asn Val Lys Ile Gln Lys Lys Pro Gly Lys Ser Glu
                85   90   95

Leu Arg Ile Asn Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met Cys
            100   105   110

Lys Val Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser Ala Asn Ile Thr
        115   120   125

Ile Val Glu Ser Asn Asp Leu Thr Thr Gly Met Ser Ala Ser Thr Glu
    130   135   140

Arg Pro Tyr Val Ser Ser Glu Ser Pro Ile Arg Ile Ser Val Ser Thr
145   150   155   160

Glu Gly Ala Asn Thr Ser Ser Ser Thr Ser Thr Ser Thr Thr Gly Thr
                165   170   175

Ser His Leu Ile Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
            180   185   190

Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
        195   200   205

Leu Cys Lys
    210

<210> 92
<211> 73
<212> PRT
<213> Mus musculus

<400> 92

Met Ser Ala Ser Thr Glu Arg Pro Tyr Val Ser Ser Glu Ser Pro Ile
1   5   10   15

Arg Ile Ser Val Ser Thr Glu Gly Ala Asn Thr Ser Ser Ser Thr Ser
            20   25   30

Thr Ser Thr Thr Gly Thr Ser His Leu Ile Lys Cys Ala Glu Lys Glu
        35   40   45

Lys Thr Phe Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys Asp Leu
    50   55   60

Ser Asn Pro Ser Arg Tyr Leu Cys Lys
65   70

<210> 93
<211> 137
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (113)..(113)
<223> X = undefined amino acid

<400> 93

```
Thr Arg Pro Lys Leu Lys Lys Met Lys Ser Gln Thr Gly Gln Val Gly
1               5                   10                  15

Glu Lys Gln Ser Leu Lys Cys Glu Ala Ala Ala Ile Asn Pro Gln Pro
            20                  25                  30

Ser Tyr Arg Trp Phe Lys Asp Gly Lys Glu Leu Asn Arg Ser Arg Asp
        35                  40                  45

Ile Arg Ile Lys Tyr Gly Asn Gly Arg Lys Asn Ser Arg Leu Gln Phe
    50                  55                  60

Asn Lys Val Lys Val Glu Asp Ala Gly Glu Tyr Val Cys Glu Ala Glu
65                  70                  75                  80

Asn Ile Leu Gly Lys Asp Thr Val Arg Gly Arg Leu Tyr Val Asn Ser
                85                  90                  95

Val Thr Thr Thr Leu Ser Ser Trp Ser Gly His Ala Gly Lys Cys Asn
            100                 105                 110

Xaa Thr Ala Lys Ser Tyr Cys Val Asn Gly Gly Val Cys Tyr Tyr Ile
        115                 120                 125

Glu Gly Ile Asn Gln Leu Ser Cys Lys
    130                 135
```

<210> 94
<211> 73
<212> PRT
<213> Homo sapiens

<400> 94

```
Ser Ser Ser Ser Phe Asp Val Gly His Glu Gly Asp Asp Ser Trp Gly
1               5                   10                  15

Leu Gly Ile Val Ser Val Arg His Trp His Met Ser Leu Ile Pro Ser
            20                  25                  30

Val Ser Thr Thr Leu Ser Ser Trp Ser Gly His Ala Arg Lys Cys Asn
```

35 40 45

Glu Thr Ala Lys Ser Tyr Cys Val Asn Gly Gly Val Cys Tyr Tyr Ile
50 55 60

Glu Gly Ile Asn Gln Leu Ser Cys Lys
65 70

<210> 95
<211> 78
<212> PRT
<213> Homo sapiens

<400> 95

Glu Ile Asn Ile Ile Ile Trp Tyr Tyr Phe Pro Ser Ala Trp Arg Thr
1 5 10 15

Cys Phe Asn Ile Ser Ser Ser Val Gly Leu Leu Leu Thr Asn Ser Tyr
20 25 30

Lys Phe Tyr Thr Thr Thr Tyr Ser Thr Glu Arg Ser Glu His Phe Lys
35 40 45

Pro Cys Arg Asp Lys Asp Leu Ala Tyr Cys Leu Asn Asp Gly Glu Cys
50 55 60

Phe Val Ile Glu Thr Leu Thr Gly Ser His Lys His Cys Arg
65 70 75

<210> 96
<211> 42
<212> PRT
<213> Homo sapiens

<400> 96

Asn Tyr Leu Gln Ile Lys Met Pro Thr Asp His Glu Glu Pro Cys Gly
1 5 10 15

Pro Ser His Lys Ser Phe Cys Leu Asn Gly Gly Leu Cys Tyr Val Ile
20 25 30

Pro Thr Ile Pro Ser Pro Phe Cys Arg Lys
35 40

<210> 97
<211> 36
<212> PRT
<213> Homo sapiens

<400> 97

Met Pro Thr Asp His Glu Glu Pro Cys Gly Pro Ser His Lys Ser Phe
1 5 10 15

Cys Leu Asn Gly Gly Leu Cys Tyr Val Ile Pro Thr Ile Pro Ser Pro
20 25 30

Phe Cys Arg Lys
35

<210> 98
<211> 36
<212> PRT
<213> Homo sapiens

<400> 98

Met Pro Thr Asp His Glu Glu Pro Cys Gly Pro Ser His Lys Ser Phe
1 5 10 15

Cys Leu Asn Gly Gly Leu Cys Tyr Val Ile Pro Thr Ile Pro Ser Pro
20 25 30

Phe Cys Arg Lys
35

<210> 99
<211> 37
<212> PRT
<213> Mus musculus

<400> 99

Met Pro Thr Gly Asn Phe Leu Ser Arg Ala Ala Leu Trp Ser Gln Ala
1 5 10 15

Gln Val Ile Leu Pro Gln Trp Gly Asp Leu Leu Cys Asp Pro Tyr Tyr
20 25 30

Pro Gln Pro Ile Leu
35

<210> 100
<211> 37
<212> PRT
<213> Mus musculus

<400> 100

Met Pro Thr Gly Asn Phe Leu Ser Arg Ala Ala Leu Trp Ser Gln Ala
1 5 10 15

Gln Val Ile Leu Pro Gln Trp Gly Asp Leu Leu Cys Asp Pro Tyr Tyr
20 25 30

Pro Gln Pro Ile Leu
35

<210> 101
<211> 25
<212> PRT
<213> Homo sapiens

<400> 101

```
Ser His Lys Ser Phe Cys Leu Asn Gly Gly Leu Cys Tyr Val Ile Pro
1               5                   10                  15

Thr Ile Pro Ser Pro Phe Cys Arg Lys
            20                  25
```

<210> 102
<211> 30
<212> PRT
<213> Sus scrofa

<400> 102

```
Glu Pro Cys Gly Pro Ser His Arg Ser Phe Cys Leu Asn Gly Gly Ile
1               5                   10                  15

Cys Tyr Val Ile Pro Thr Ile Pro Ser Pro Phe Cys Arg Lys
            20                  25                  30
```

<210> 103
<211> 30
<212> PRT
<213> Sus scrofa

<400> 103

```
Glu Pro Cys Gly Pro Ser His Arg Ser Phe Cys Leu Asn Gly Gly Ile
1               5                   10                  15

Cys Tyr Val Ile Pro Thr Ile Pro Ser Pro Phe Cys Arg Lys
            20                  25                  30
```

<210> 104
<211> 46
<212> PRT
<213> Mus musculus

<400> 104

```
Cys Leu Phe Ala Pro Ala Asp Ser Pro Val Ala Ala Ala Val Val Ser
1               5                   10                  15

His Phe Asn Lys Cys Pro Asp Ser His Thr Gln Tyr Cys Phe His Gly
            20                  25                  30

Thr Cys Arg Phe Leu Val Gln Glu Glu Lys Pro Ala Cys Val
        35                  40                  45
```

<210> 105

<211> 51
<212> PRT
<213> Homo sapiens

<400> 105

Asp Leu Ser Pro Ala Ser Phe Leu Ser Pro Ala Asp Pro Pro Val Ala
1   5   10   15

Ala Ala Val Val Ser His Phe Asn Asp Cys Pro Asp Ser His Thr Gln
            20   25   30

Phe Cys Phe His Gly Thr Cys Arg Phe Leu Val Gln Glu Asp Lys Pro
        35   40   45

Ala Cys Val
    50

<210> 106
<211> 42
<212> PRT
<213> Homo sapiens

<400> 106

Val Gln Thr Glu Asp Asn Pro Arg Val Ala Gln Val Ser Ile Thr Lys
1   5   10   15

Cys Ser Ser Asp Met Asn Gly Tyr Cys Leu His Gly Gln Cys Ile Tyr
            20   25   30

Leu Val Asp Met Ser Gln Asn Tyr Cys Arg
        35   40

<210> 107
<211> 40
<212> PRT
<213> Homo sapiens

<400> 107

Gln Thr Glu Asp Asn Pro Arg Val Ala Gln Val Ser Ile Thr Lys Cys
1   5   10   15

Ser Ser Asp Met Asn Gly Tyr Cys Leu His Gly Gln Cys Ile Tyr Leu
            20   25   30

Val Asp Met Ser Gln Asn Tyr Cys
        35   40

<210> 108
<211> 42
<212> PRT
<213> Mus musculus

<400> 108

Val Gln Met Glu Asp Asp Pro Arg Val Ala Gln Val Gln Ile Thr Lys
1   5   10   15

Cys Ser Ser Asp Met Asp Gly Tyr Cys Leu His Gly Gln Cys Ile Tyr
            20   25   30

Leu Val Asp Met Arg Glu Lys Phe Cys Arg
        35   40

<210> 109
<211> 93
<212> PRT
<213> Homo sapiens

<400> 109

Met Thr Ala Gly Arg Arg Met Glu Met Leu Cys Ala Gly Arg Val Pro
1   5   10   15

Ala Leu Leu Leu Cys Leu Gly Phe His Leu Leu Gln Ala Val Leu Ser
            20   25   30

Thr Thr Val Ile Pro Ser Cys Ile Pro Gly Glu Ser Ser Asp Asn Cys
        35   40   45

Thr Ala Leu Val Gln Thr Glu Asp Asn Pro Arg Val Ala Gln Val Ser
    50   55   60

Ile Thr Lys Cys Ser Ser Asp Met Asn Gly Tyr Cys Leu His Gly Gln
65   70   75   80

Cys Ile Tyr Leu Val Asp Met Ser Gln Asn Tyr Cys Arg
                85   90

<210> 110
<211> 93
<212> PRT
<213> Homo sapiens

<400> 110

Met Thr Ala Gly Arg Arg Met Glu Met Leu Cys Ala Gly Arg Val Pro
1   5   10   15

Ala Leu Leu Leu Cys Leu Gly Phe His Leu Leu Gln Ala Val Leu Ser
            20   25   30

Thr Thr Val Ile Pro Ser Cys Ile Pro Gly Glu Ser Ser Asp Asn Cys
        35   40   45

Thr Ala Leu Val Gln Thr Glu Asp Asn Pro Arg Val Ala Gln Val Ser
    50   55   60

Ile Thr Lys Cys Ser Ser Asp Met Asn Gly Tyr Cys Leu His Gly Gln
65 70 75 80

Cys Ile Tyr Leu Val Asp Met Ser Gln Asn Tyr Cys Arg
85 90

<210> 111
<211> 180
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> X = undefined amino acid

<220>
<221> misc_feature
<222> (118)..(118)
<223> X = undefined amino acid

<400> 111

Pro Gly Glu Lys Ala Thr Arg Pro Lys Leu Lys Lys Met Lys Ser Gln
1 5 10 15

Thr Gly Gln Val Gly Glu Lys Gln Ser Leu Lys Cys Glu Ala Ala Ala
20 25 30

Gly Asn Pro Gln Pro Ser Tyr Arg Trp Phe Lys Asp Gly Lys Glu Leu
35 40 45

Asn Arg Ser Arg Asp Ile Arg Ile Lys Tyr Gly Asn Gly Arg Lys Asn
50 55 60

Ser Arg Leu Gln Phe Asn Lys Val Lys Val Glu Asp Ala Gly Glu Tyr
65 70 75 80

Val Cys Glu Ala Glu Asn Ile Leu Gly Lys Asp Thr Val Gly Gly Arg
85 90 95

Leu Tyr Val Asn Ser Val Thr Thr Thr Leu Ser Ser Trp Ser Gly His
100 105 110

Ala Arg Lys Cys Asn Xaa Thr Ala Lys Ser Tyr Cys Val Asn Gly Gly
115 120 125

Val Cys Tyr Tyr Ile Glu Gly Ile Asn Gln Leu Ser Cys Lys Ala Pro
130 135 140

Gly Leu His Cys Leu Glu Leu Gly Thr Gln Ser His His Phe Pro Ile
145 150 155 160

Ser Ala Ser Pro Gly Ser Ser Gln Gly Ser Trp Asn Gln Leu Pro Gln
165 170 175

His Pro Leu Ser
180

<210> 112
<211> 120
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (13)..(13)
<223> X = undefined amino acid

<400> 112

Glu Ala Glu Asn Ile Leu Gly Lys Asp Thr Val Arg Xaa Arg Leu Tyr
1 5 10 15

Val Asn Ser Val Ser Thr Thr Leu Ser Ser Trp Ser Gly His Ala Arg
20 25 30

Lys Cys Asn Glu Thr Ala Lys Ser Tyr Cys Val Asn Gly Gly Val Cys
35 40 45

Tyr Tyr Ile Glu Gly Ile Asn Gln Leu Ser Cys Lys Ala His Gly Leu
50 55 60

His Cys Leu Glu Leu Gly Thr Gln Ser His His Phe Pro Ile Ser Ala
65 70 75 80

Ser Pro Gly Ser Ser Gln Gly Ser Trp Asn Gln Leu Pro Gln His Pro
85 90 95

Leu Ser Ala Leu Gly Gly Glu Gly Ser Pro Gly Gly Asp Ala Val Arg
100 105 110

Thr Pro Gly Pro Gln Ser Cys Ala
115 120

<210> 113
<211> 76
<212> PRT
<213> Mus musculus

<400> 113

Val Arg Gln Arg Arg Glu Thr Pro Ser Pro Pro Ile Ala Gly Ser Arg
1 5 10 15

Met Ala Arg Asn Ser Thr Gly Val Val Ile Phe Ala Ser Ser Met Ala

20 25 30

Met Ala Val Ser Thr Thr Leu Ser Ser Trp Ser Gly His Ala Arg Lys
35 40 45

Cys Asn Glu Thr Ala Lys Ser Tyr Cys Val Asn Gly Gly Val Cys Tyr
50 55 60

Tyr Ile Glu Gly Ile Asn Gln Leu Ser Cys Lys Gly
65 70 75

<210> 114
<211> 167
<212> PRT
<213> Danio rerio

<400> 114

Lys Asp Cys Ala Ser Ala Pro Lys Val Lys Pro Met Asp Ser Gln Trp
1 5 10 15

Leu Gln Glu Gly Lys Lys Leu Thr Leu Lys Cys Glu Ala Val Gly Asn
20 25 30

Pro Ser Pro Ser Phe Asn Trp Tyr Lys Asp Gly Ser Gln Leu Arg Gln
35 40 45

Lys Lys Thr Val Lys Ile Lys Thr Asn Lys Lys Asn Ser Lys Leu His
50 55 60

Ile Ser Lys Val Arg Leu Glu Asp Ser Gly Asn Tyr Thr Cys Val Val
65 70 75 80

Glu Asn Ser Leu Gly Arg Glu Asn Ala Thr Ser Phe Val Ser Val Gln
85 90 95

Ser Ile Thr Thr Thr Leu Ser Pro Gly Ser Ser His Ala Arg Lys Cys
100 105 110

Asn Glu Thr Glu Lys Thr Tyr Cys Ile Asn Gly Gly Asp Cys Tyr Phe
115 120 125

Ile His Gly Ile Asn Gln Leu Ser Cys Lys Cys Pro Asn Asp Tyr Thr
130 135 140

Gly Glu Arg Cys Gln Thr Ser Val Met Ala Gly Phe Tyr Lys Ala Glu
145 150 155 160

Glu Leu Tyr Gln Asn Glu Cys
165

<210> 115
<211> 84
<212> PRT
<213> Gallus gallus

<400> 115

Ala Val Gln Ser Leu Glu Leu Leu Gln Gln Thr Trp Arg Leu Ser Thr
1 5 10 15

Leu Gln Phe Glu Tyr Asp Arg Arg Val Ala Cys Gly Phe His Tyr Thr
20 25 30

Thr Thr Tyr Ser Thr Glu Arg Ser Glu His Phe Lys Pro Cys Lys Asp
35 40 45

Lys Asp Leu Ala Tyr Cys Leu Asn Glu Gly Glu Cys Phe Val Ile Glu
50 55 60

Thr Leu Thr Gly Ser His Lys His Cys Arg Ser Asn Cys Pro Ser Gly
65 70 75 80

Val Phe Cys Trp

<210> 116
<211> 77
<212> PRT
<213> Gallus gallus

<400> 116

Met Arg Thr Asp His Glu Glu Leu Cys Gly Thr Ser Tyr Gly Ser Phe
1 5 10 15

Cys Leu Asn Gly Gly Ile Cys Tyr Met Ile Pro Thr Val Pro Ser Pro
20 25 30

Phe Cys Arg His Leu Pro Lys Ala Ala Asn Gln Ala Ser Ala Leu His
35 40 45

Lys Ser Val Phe Ser Ile Phe Val Leu His Thr Asp Thr Thr Ala Leu
50 55 60

Pro Ser Cys His Leu Met Pro Ala His Phe Tyr Thr Gln
65 70 75

<210> 117
<211> 65
<212> PRT
<213> Mus musculus

<400> 117

Met Pro Thr Asp His Glu Gln Pro Cys Gly Pro Arg His Arg Ser Phe

1 5 10 15

Cys Leu Asn Gly Gly Ile Cys Ile Asp Pro Tyr Tyr Pro His Pro Phe
20 25 30

Cys Arg Phe Tyr His Leu Phe Leu Arg His Cys Leu Leu Lys Pro Phe
35 40 45

Val Gln Leu Gly Thr Leu Val Tyr Pro Val Phe Leu Lys Glu Leu Phe
50 55 60

His
65

<210> 118
<211> 70
<212> PRT
<213> Homo sapiens

<400> 118

Asp Val Ile Ala Gln His Lys Pro Glu Ser Glu Asn Thr Ser Asp Lys
1 5 10 15

Pro Lys Arg Lys Lys Lys Gly Gly Lys Asn Gly Lys Asn Arg Arg Asn
20 25 30

Arg Lys Lys Lys Asn Pro Cys Asp Ala Glu Phe Gln Asn Phe Cys Ile
35 40 45

His Gly Glu Cys Lys Tyr Ile Glu His Leu Glu Ala Val Thr Cys Asn
50 55 60

Val Ser Arg Ile Phe Pro
65 70

<210> 119
<211> 112
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (2)..(2)
<223> X = undefined amino acid

<400> 119

Leu Xaa Ala Thr Thr Gln Ser Lys Trp Lys Gly His Ser Ser Arg Cys
1 5 10 15

Pro Lys Gln Tyr Lys His Tyr Cys Ile Lys Gly Arg Cys Arg Phe Val
20 25 30

Val Ala Glu Gln Thr Pro Ser Cys Val Pro Leu Arg Lys Arg Arg Lys
35 40 45

Arg Lys Lys Lys Glu Glu Glu Met Glu Thr Leu Gly Lys Asp Met Thr
50 55 60

Pro Ile Asn Glu Asp Ile Glu Glu Thr Asn Ile Ala Tyr Lys Ala Met
65 70 75 80

Lys Leu Pro Pro Gly Trp Trp Gln Ala Ala Lys Cys Leu Ala His Leu
85 90 95

Lys Met Asp Arg Met Arg Leu Arg Lys Thr Ala Ser Arg His Glu Phe
100 105 110

<210> 120
<211> 119
<212> PRT
<213> Mus musculus

<400> 120

Lys Ser Leu Thr Trp Lys Ser Phe Asn Phe Leu Ser Leu Leu Leu Pro
1 5 10 15

Leu Gly Ser Thr Gly Thr Arg Arg Ile Leu Cys Pro Leu Ser Thr Pro
20 25 30

Ser Cys Ser Ala Gly Leu Ala Ile Leu His Cys Val Val Ala Asp Gly
35 40 45

Asn Thr Thr Arg Thr Pro Glu Thr Asn Gly Ser Leu Cys Gly Ala Pro
50 55 60

Gly Glu Asn Cys Thr Gly Thr Thr Pro Arg Gln Lys Val Lys Thr His
65 70 75 80

Phe Ser Arg Cys Pro Lys Gln Tyr Lys His Tyr Cys Ile His Gly Arg
85 90 95

Cys Arg Phe Val Val Asp Glu Gln Thr Pro Ser Cys Met Ala Arg Leu
100 105 110

Ser Ile Tyr Leu Trp Arg Asn
115

<210> 121
<211> 141
<212> PRT
<213> Cercopithecus aethiops (African green monkey)

<400> 121

Met Lys Leu Leu Pro Ser Val Val Leu Lys Leu Leu Leu Ala Ala Val
1   5   10   15

Leu Ser Ala Leu Val Thr Gly Glu Ser Leu Glu Gln Leu Arg Arg Gly
            20   25   30

Pro Ala Ala Gly Thr Ser Asn Pro Asp Pro Ser Thr Gly Ser Thr Asp
        35   40   45

Gln Leu Leu Arg Leu Gly Gly Gly Arg Asp Arg Lys Val Arg Asp Leu
    50   55   60

Gln Glu Ala Asp Leu Asp Leu Leu Arg Val Thr Leu Ser Ser Lys Pro
65   70   75   80

Gln Ala Leu Ala Thr Pro Ser Lys Glu Glu His Gly Lys Arg Lys Lys
                85   90   95

Lys Gly Lys Gly Leu Gly Lys Lys Arg Asp Pro Cys Leu Arg Lys Tyr
            100   105   110

Lys Asp Phe Cys Ile His Gly Glu Cys Lys Tyr Val Lys Glu Leu Arg
        115   120   125

Ala Pro Ser Cys Met Ala Ala Gly Gln Lys Asp Val Thr
    130   135   140

<210> 122
<211> 79
<212> PRT
<213> Homo sapiens

<400> 122

Met Thr Ala Leu Thr Glu Glu Ala Ala Val Thr Val Thr Pro Pro Ile
1   5   10   15

Thr Ala Gln Gln Ala Asp Asn Ile Glu Gly Pro Ile Ala Leu Lys Phe
            20   25   30

Ser His Leu Cys Leu Glu Asp His Asn Ser Tyr Cys Ile Asn Gly Ala
        35   40   45

Cys Ala Phe His His Glu Leu Glu Lys Ala Ile Cys Arg Cys Leu Lys
    50   55   60

Leu Lys Ser Pro Tyr Asn Val Cys Ser Gly Glu Arg Arg Pro Leu
65   70   75

<210> 123
<211> 96

<212> PRT
<213> Homo sapiens

<400> 123

Gly Thr Arg Glu Ala Leu Cys Tyr Arg Cys Phe Cys Pro Leu Asn Thr
1 5 10 15

Ala Met Arg Ala Leu Thr Glu Glu Ala Ala Val Thr Val Thr Pro Pro
20 25 30

Ile Thr Ala Gln Gln Ala Asp Asn Ile Glu Gly Pro Ile Ala Leu Lys
35 40 45

Phe Ser His Leu Cys Leu Glu Asp His Asn Ser Tyr Cys Ile Asn Gly
50 55 60

Ala Cys Ala Phe His His Glu Leu Glu Lys Ala Ile Cys Arg Cys Leu
65 70 75 80

Lys Leu Lys Ser Pro Tyr Asn Val Cys Ser Gly Glu Arg Arg Pro Leu
85 90 95

<210> 124
<211> 96
<212> PRT
<213> Homo sapiens

<400> 124

Gly Thr Arg Glu Ala Leu Cys Tyr Arg Cys Phe Cys Pro Leu Asn Thr
1 5 10 15

Ala Met Arg Ala Leu Thr Glu Glu Ala Ala Val Thr Val Thr Pro Pro
20 25 30

Ile Thr Ala Gln Gln Ala Asp Asn Ile Glu Gly Pro Ile Ala Leu Lys
35 40 45

Phe Ser His Leu Cys Leu Glu Asp His Asn Ser Tyr Cys Ile Asn Gly
50 55 60

Ala Cys Ala Phe His His Glu Leu Glu Lys Ala Ile Cys Arg Cys Leu
65 70 75 80

Lys Leu Lys Ser Pro Tyr Asn Val Cys Ser Gly Glu Arg Arg Pro Leu
85 90 95

<210> 125
<211> 97
<212> PRT
<213> Homo sapiens

<400> 125

Leu Gln Glu Met Ala Leu Gly Val Pro Ile Ser Val Tyr Leu Leu Phe
1   5   10   15

Asn Ala Met Thr Ala Leu Thr Glu Glu Ala Ala Val Thr Val Thr Pro
            20   25   30

Pro Ile Thr Ala Gln Gln Ala Asp Asn Ile Glu Gly Pro Ile Ala Leu
        35   40   45

Lys Phe Ser His Leu Cys Leu Glu Asp His Asn Ser Tyr Cys Ile Asn
    50   55   60

Gly Ala Cys Ala Phe His His Glu Leu Glu Lys Ala Ile Cys Arg Cys
65   70   75   80

Leu Lys Leu Lys Ser Pro Tyr Asn Val Cys Ser Gly Glu Arg Arg Pro
                85   90   95

Leu

<210> 126
<211> 115
<212> PRT
<213> Homo sapiens

<400> 126

Lys Asp Lys Arg Lys Lys Val Lys Gln Leu Gln Glu Met Ala Leu Gly
1   5   10   15

Val Pro Ile Ser Val Tyr Leu Leu Phe Asn Ala Met Thr Ala Leu Thr
            20   25   30

Glu Glu Ala Ala Val Thr Val Thr Pro Pro Ile Thr Ala Gln Gln Gly
        35   40   45

Asn Trp Thr Val Asn Lys Thr Glu Ala Asp Asn Ile Glu Gly Pro Ile
    50   55   60

Ala Leu Lys Phe Ser His Leu Cys Leu Glu Asp His Asn Ser Tyr Cys
65   70   75   80

Ile Asn Gly Ala Cys Ala Phe His His Glu Leu Glu Lys Ala Ile Cys
                85   90   95

Arg Cys Leu Lys Leu Lys Ser Pro Tyr Asn Val Cys Ser Gly Glu Arg
            100   105   110

Arg Pro Leu
        115

<210> 127
<211> 94
<212> PRT
<213> Homo sapiens

<400> 127

Met Ala Leu Gly Val Pro Ile Ser Val Tyr Leu Leu Phe Asn Ala Met
1 5 10 15

Thr Ala Leu Thr Glu Glu Ala Ala Val Thr Val Thr Pro Pro Ile Thr
20 25 30

Ala Gln Gln Ala Asp Asn Ile Glu Gly Pro Ile Ala Leu Lys Phe Ser
35 40 45

His Leu Cys Leu Glu Asp His Asn Ser Tyr Cys Ile Asn Gly Ala Cys
50 55 60

Ala Phe His His Glu Leu Glu Lys Ala Ile Cys Arg Cys Leu Lys Leu
65 70 75 80

Lys Ser Pro Tyr Asn Val Cys Ser Gly Glu Arg Arg Pro Leu
85 90

<210> 128
<211> 117
<212> DNA
<213> Homo sapiens

<400> 128
actgggacaa gccatcttgt aaaatgtgcg gagaaggaga aaactttctg tgtgaatgga
60

ggggagtgct tcatggtgaa agacctttca aacccctcga gatacttgtg caagtaa
117

<210> 129
<211> 108
<212> DNA
<213> Homo sapiens

<400> 129
tcctggtcgg ggcacgcccg gaagtgcaac gagacagcca agtcctattg cgtcaatgga
60

ggcgtctgct actacatcga gggcatcaac cagctctcct gcaagtaa
108

<210> 130
<211> 114
<212> DNA
<213> Homo sapiens

<400> 130
gagcgatccg agcacttcaa accctgccga gacaaggacc ttgcatactg tctcaatgat

60

ggcgagtgct ttgtgatcga aaccctgacc ggatcccata aacactgtcg gtaa
114

<210> 131
<211> 99
<212> DNA
<213> Homo sapiens

<400> 131
gatcacgaag agccctgtgg tcccagtcac aagtcgtttt gcctgaatgg ggggctttgt
60

tatgtgatac ctactattcc cagcccattt tgtaggtga
99

<210> 132
<211> 108
<212> DNA
<213> Homo sapiens

<400> 132
tccgtaagaa atagtgactc tgaatgtccc ctgtcccacg atgggtactg cctccatgat
60

ggtgtgtgca tgtatattga agcattggac aagtatgcat gcaagtaa
108

<210> 133
<211> 105
<212> DNA
<213> Homo sapiens

<400> 133
gcagtggtgt cccattttaa tgactgccca gattccccaca ctcagttctg cttccatgga
60

acctgcaggt ttttggtgca ggaggacaag ccagcatgtg tgtaa
105

<210> 134
<211> 105
<212> DNA
<213> Homo sapiens

<400> 134
aagcggaaag gccacttctc taggtgcccc aagcaataca agcattactg catcaaaggg
60

agatgccgct tcgtggtggc cgagcagacg ccctcctgtg tgtaa
105

<210> 135
<211> 105
<212> DNA
<213> Homo sapiens

<400> 135
agaaacagaa agaagaaaaa tccatgtaat gcagaatttc aaaatttctg cattcacgga
60

gaatgcaaat atatagagca cctggaagca gtaacatgca agtaa
105

<210> 136
<211> 105
<212> DNA
<213> Homo sapiens

<400> 136
gggctaggga agaagaggga cccatgtctt cggaaataca aggacttctg catccatgga
60

gaatgcaaat atgtgaagga gctccgggct ccctcctgca tgtaa
105

<210> 137
<211> 105
<212> DNA
<213> Homo sapiens

<400> 137
gtggctcaag tgtcaataac aaagtgtagc tctgacatga atggctattg tttgcatgga
60

cagtgcatct atctggtgga catgagtcaa aactactgca ggtaa
105

<210> 138
<211> 105
<212> DNA
<213> Mus musculus

<400> 138
gtagctctga agttctctca tccttgtctg gaagaccata atagttactg cattaatgga
60

gcatgtgcat tccaccatga gctgaagcaa gccatttgca ggtaa
105

<210> 139
<211> 105
<212> DNA
<213> Homo sapiens

<400> 139
atagccttga agttctcaca cctttgcctg gaagatcata acagttactg catcaacggt
60

gcttgtgcat tccaccatga gctagagaaa gccatctgca ggtaa
105

<210> 140
<211> 1651
<212> DNA
<213> Homo sapiens

<400> 140
acggcacgag gagccggcga ggagttcccc gaaacttgtt ggaactccgg gctcgcgcgg
60

aggccaggag ctgagcggcg gcggctgccg gacgatggga gcgtgagcag gacggtgata
120

acctctcccc gatcgggttg cgagggcgcc gggcagaggc caggacgcga gccgccagcg
180

gcgggaccca tcgacgactt cccggggcga caggagcagc cccgagagcc agggcgagcg
240

cccgttccag gtggccggac cgcccgccgc gtccgcgccg cgctccctgc aggcaacggg
300

agacgccccc gcgcagcgcg agcgcctcag cgcggccgct cgctctcccc atcgagggac
360

aaacttttcc caaacccgat ccgagccctt ggaccaaact cgcctgcgcc gagagccgtc
420

cgcgtagagc gctccgtctc cggcgagatg tccgagcgca aagaaggcag aggcaaaggg
480

aagggcaaga agaaggagcg aggctccggc aagaagccgg agtccgcggc gggcagccag
540

agcccagcct tgcctcccca attgaaagag atgaaaagcc aggaatcggc tgcaggttcc
600

aaactagtcc ttcggtgtga aaccagttct gaatactcct ctctcagatt caagtggttc
660

aagaatggga atgaattgaa tcgaaaaaac aaaccacaaa atatcaagat acaaaaaaag
720

ccagggaagt cagaacttcg cattaacaaa gcatcactgg ctgattctgg agagtatatg
780

tgcaaagtga tcagcaaatt aggaaatgac agtgcctctg ccaatatcac catcgtggaa
840

tcaaacgaga tcatcactgg tatgccagcc tcaactgaag gagcatatgt gtcttcagag
900

tctcccatta gaatatcagt atccacagaa ggagcaaata cttcttcatc tacatctaca
960

tccaccactg ggacaagcca tcttgtaaaa tgtgcggaga aggagaaaac tttctgtgtg
1020

aatggagggg agtgcttcat ggtgaaagac ctttcaaacc cctcgagata cttgtgcaag
1080

taagaaaaga aatcctgtgt gtcgcttatg tctataactc cttgtttcag atgattctat
1140

gtctcatgat tgattgttgc ttttttttcca attttgttgc atcatgttga ataatgctgt
1200

tttatatgta gagtctttta aaacattcac accattcgtc atcactcctc tgtcatatgc
1260

agttttgttt tttgctcttt tcaatgtgtg tgaggtgttt tttgtttttg tttttgtttt
1320

tttgccatgt tatttatagt gttgctttcc ttgtgctttc cttgtggttt tcttggttgg
1380

ttattcagaa aagatgtgca gatatcacag aggcctatag ccttttggta tctacttcta
1440

catccaatgt atgaattaag ctgtaagata atgttgcttt cttatcccag tgatcacctg
1500

ccaaatgaat aagacaacaa agagaagcag aagggcaaga agattattta ctgacatata
1560

tctattacac ttgggattgt gcttactgtt gcataactat tttttaaacg gagtttagtt
1620

ttatattgct agtaaaaaaa aaaaaaaaaa a
1651

<210> 141
<211> 675
<212> DNA
<213> Homo sapiens

<400> 141
ctacatctac atccaccact gggacaagcc atcttgtaaa atgtgcggag aaggagaaaa
60

ctttctgtgt gaatggaggg gagtgcttca tggtgaaaga cctttcaaac ccctcgagat
120

acttgtgcaa gtaagaaaag aaatcctgtg tgtcgcttat gtctataact ccttgtttca
180

gatgattcta tgtctcatga tgtattgttg ctttttttcc aattttgttg catcatgttg
240

aataatgctg ttttatatgt agagtgtttt aaaacattca caccattcgt catcactcct
300

ctgtcatatg cagaattgtt ttttgctctt ttcaatgtgt gtgaggtgtt ttttgttttt
360

gtttttgttt tttgccatgt tatttatagt gttgctttcc ttgtggtttt tcttgttgtt
420

attcagaaaa gatgtgcaga tatcacagag gcctataact tttggtatct acttctacat
480

ccaatgtatg aattaagctg taagataatg ttgctttctt atcccrgtga tcacctgcca
540

aatgaataag acaacaaaga gaagcagaag ggcagaagat tatttactga catatatcta
600

ttacacttgg gattgtctya ctgttgcata actatttttt aaacggagtt tagttttata
660

ttgctagtaa aaaaa
675

<210> 142
<211> 675
<212> DNA
<213> Homo sapiens

<400> 142
ctacatctac atccaccact gggacaagcc atcttgtaaa atgtgcggag aaggagaaaa

60

ctttctgtgt gaatggaggg gagtgcttca tggtgaaaga cctttcaaac ccctcgagat
120

acttgtgcaa gtaagaaaag aaatcctgtg tgtcgcttat gtctataact ccttgtttca
180

gatgattcta tgtctcatga tgtattgttg ctttttttcc aattttgttg catcatgttg
240

aataatgctg ttttatatgt agagtgtttt aaaacattca caccattcgt catcactcct
300

ctgtcatatg cagaattgtt ttttgctctt ttcaatgtgt gtgaggtgtt ttttgttttt
360

gtttttgttt tttgccatgt tatttatagt gttgctttcc ttgtggtttt tcttgttgtt
420

attcagaaaa gatgtgcaga tatcacagag gcctataact tttggtatct acttctacat
480

ccaatgtatg aattaagctg taagataatg ttgctttctt atcccrgtga tcacctgcca
540

aatgaataag acaacaaaga gaagcagaag ggcagaagat tatttactga catatatcta
600

ttacacttgg gattgtctya ctgttgcata actatttttt aaacggagtt tagttttata
660

ttgctagtaa aaaaa
675

<210> 143
<211> 1651
<212> DNA
<213> Homo sapiens

<400> 143
acggcacgag gagccggcga ggagttcccc gaaacttgtt ggaactccgg gctcgcgcgg
60

aggccaggag ctgagcggcg gcggctgccg gacgatggga gcgtgagcag gacggtgata
120

acctctcccc gatcgggttg cgagggcgcc gggcagaggc caggacgcga gccgccagcg
180

gcgggaccca tcgacgactt cccggggcga caggagcagc cccgagagcc agggcgagcg
240

cccgttccag gtggccggac cgcccgccgc gtccgcgccg cgctccctgc aggcaacggg
300

agacgccccc gcgcagcgcg agcgcctcag cgcggccgct cgctctcccc atcgagggac
360

aaacttttcc caaacccgat ccgagccctt ggaccaaact cgcctgcgcc gagagccgtc
420

cgcgtagagc gctccgtctc cggcgagatg tccgagcgca aagaaggcag aggcaaaggg
480

aagggcaaga agaaggagcg aggctccggc aagaagccgg agtccgcggc gggcagccag 540

agcccagcct tgcctcccca attgaaagag atgaaaagcc aggaatcggc tgcaggttcc 600

aaactagtcc ttcggtgtga aaccagttct gaatactcct ctctcagatt caagtggttc 660

aagaatggga atgaattgaa tcgaaaaaac aaaccacaaa atatcaagat acaaaaaaag 720

ccagggaagt cagaacttcg cattaacaaa gcatcactgg ctgattctgg agagtatatg 780

tgcaaagtga tcagcaaatt aggaaatgac agtgcctctg ccaatatcac catcgtggaa 840

tcaaacgaga tcatcactgg tatgccagcc tcaactgaag gagcatatgt gtcttcagag 900

tctcccatta gaatatcagt atccacagaa ggagcaaata cttcttcatc tacatctaca 960

tccaccactg ggacaagcca tcttgtaaaa tgtgcggaga aggagaaaac tttctgtgtg 1020

aatggagggg agtgcttcat ggtgaaagac ctttcaaacc cctcgagata cttgtgcaag 1080

taagaaaaga aatcctgtgt gtcgcttatg tctataactc cttgtttcag atgattctat 1140

gtctcatgat tgattgttgc ttttttttcca attttgttgc atcatgttga ataatgctgt 1200

tttatatgta gagtctttta aaacattcac accattcgtc atcactcctc tgtcatatgc 1260

agttttgttt tttgctcttt tcaatgtgtg tgaggtgttt tttgtttttg tttttgtttt 1320

tttgccatgt tatttatagt gttgctttcc ttgtgctttc cttgtggttt tcttggttgg 1380

ttattcagaa aagatgtgca gatatcacag aggcctatag ccttttggta tctacttcta 1440

catccaatgt atgaattaag ctgtaagata atgttgcttt cttatcccag tgatcacctg 1500

ccaaatgaat aagacaacaa agagaagcag aagggcaaga agattattta ctgacatata 1560

tctattacac ttgggattgt gcttactgtt gcataactat tttttaaacg gagtttagtt 1620

ttatattgct agtaaaaaaa aaaaaaaaaa a 1651

<210> 144
<211> 1651
<212> DNA
<213> Homo sapiens

<400> 144
acggcacgag gagccggcga ggagttcccc gaaacttgtt ggaactccgg gctcgcgcgg 60

aggccaggag ctgagcggcg gcggctgccg gacgatggga gcgtgagcag gacggtgata 120

acctctcccc gatcgggttg cgagggcgcc gggcagaggc caggacgcga gccgccagcg 180

gcgggaccca tcgacgactt cccggggcga caggagcagc cccgagagcc agggcgagcg 240

cccgttccag gtggccggac cgcccgccgc gtccgcgccg cgctccctgc aggcaacggg 300

agacgccccc gcgcagcgcg agcgcctcag cgcggccgct cgctctcccc atcgagggac 360

aaacttttcc caaacccgat ccgagccctt ggaccaaact cgcctgcgcc gagagccgtc 420

cgcgtagagc gctccgtctc cggcgagatg tccgagcgca aagaaggcag aggcaaaggg 480

aagggcaaga agaaggagcg aggctccggc aagaagccgg agtccgcggc gggcagccag 540

agcccagcct tgcctcccca attgaaagag atgaaaagcc aggaatcggc tgcaggttcc 600

aaactagtcc ttcggtgtga aaccagttct gaatactcct ctctcagatt caagtggttc 660

aagaatggga atgaattgaa tcgaaaaaac aaaccacaaa atatcaagat acaaaaaaag 720

ccagggaagt cagaacttcg cattaacaaa gcatcactgg ctgattctgg agagtatatg 780

tgcaaagtga tcagcaaatt aggaaatgac agtgcctctg ccaatatcac catcgtggaa 840

tcaaacgaga tcatcactgg tatgccagcc tcaactgaag gagcatatgt gtcttcagag 900

tctcccatta gaatatcagt atccacagaa ggagcaaata cttcttcatc tacatctaca 960

tccaccactg ggacaagcca tcttgtaaaa tgtgcggaga aggagaaaac tttctgtgtg 1020

aatggagggg agtgcttcat ggtgaaagac ctttcaaacc cctcgagata cttgtgcaag 1080

taagaaaaga aatcctgtgt gtcgcttatg tctataactc cttgtttcag atgattctat 1140

gtctcatgat tgattgttgc ttttttttcca attttgttgc atcatgttga ataatgctgt 1200

tttatatgta gagtctttta aaacattcac accattcgtc atcactcctc tgtcatatgc 1260

agttttgttt tttgctcttt tcaatgtgtg tgaggtgttt tttgtttttg tttttgtttt 1320

tttgccatgt tatttatagt gttgctttcc ttgtgctttc cttgtggttt tcttggttgg
1380

ttattcagaa aagatgtgca gatatcacag aggcctatag ccttttggta tctacttcta
1440

catccaatgt atgaattaag ctgtaagata atgttgcttt cttatcccag tgatcacctg
1500

ccaaatgaat aagacaacaa agagaagcag aagggcaaga agattattta ctgacatata
1560

tctattacac ttgggattgt gcttactgtt gcataactat tttttaaacg gagtttagtt
1620

ttatattgct agtaaaaaaa aaaaaaaaaa a
1651

<210> 145
<211> 1590
<212> DNA
<213> Mus musculus

<400> 145
gactccgggc cgcgccggca gcaggagcgg aacgcagcgc agcggcggca gctgccagga
60

gatgcgagca tagaccggac tgtgagcacc tttccctctt cgggctgtaa gggagcgaga
120

cagccaccgg agcgaggcca ctccagagcc ggcagcggca ggacccggga cacaagagta
180

gccccgagac acccccagac gtagcgggcg ctccaggtga tcgagtccac gccgctccct
240

gcaggcgaca ggcgacgccc ccgcgcagcc cggccactgg ctcttccctc ccgggacaaa
300

cttttctgca agcccttgga ccaaacttgt cgcgcgtcac cgtcgcccag ccgggtccgc
360

gtagagcgct catctttagc gagatgtctg agcgcaaaga aggcagaggc aaggggaagg
420

gcaagaagaa ggaccgggga tcccgcggga agcccgcgcc cgccgaaggc gacccgagcc
480

cagcattgcc tcccagattg aaagagatga aaagccagga gtcagctgca ggctccaagc
540

tcgtgcttcg gtgtgaaacc agctctgagt actcctcact cagattcaaa tggttcaaga
600

acgggaatga gctgaaccgt aggaataaac cacaaaacgt caagatacag aagaagccag
660

ggaagtcaga gcttcgaatc aacaaagcgt ccctggctga ctctggagaa tatatgtgca
720

aagtgatcag caagttagga aacgacagtg cctctgccaa catcaccatt gttgagtcaa
780

acgacctcac cactggcatg tcagcctcaa ctgaaagacc ttatgtgtcc tcagagtctc

840

ccattagaat atcagtttca acagaaggcg caaatacttc ttcatccaca tctacatcca
900

cgactgggac cagccatctc ataaagtgtg cggagaagga gaaaactttc tgtgtgaatg
960

gaggcgagtg cttcatggtg aaggacctgt caaacccctc aagatacttg tgcaagtaag
1020

aaatgaattc ctctctgtgc ctcgtacctg taacagctta tcccagattg ttctgtgtcg
1080

ccatgaaccc ctggcttttt tttccttact ttgttacatc ttgttttaaa taattctcat
1140

ttatttgtgg agggtttttt gaaatatttg caccatctgc cattgcctct gtcatgttca
1200

gaattgattt tacttttcaa ggttttaggg tgtttttggt tcttgatggg ttgagtattt
1260

tttttgtttg gttggttttg ggtttttgct gttttgtttt gtttttttgtt tttgttttct
1320

tttttgcctt catatatata attttgcttt cctcctggtg ttccttaata gctactgaaa
1380

gaagtgtgca aatattgtag aaagctgtca ctttgaatcc ctactttttt atcccatgta
1440

ttaattgagc cataaggtac ataaggtaac ttttttttaa cctcagtgct tacctgcaag
1500

gtgaacagga caaatagagg ttgcaagaga gcagaaagtt acctgctaaa gcatttctta
1560

tgctctggat tatggtattg ccccataatt
1590

<210> 146
<211> 1630
<212> DNA
<213> Mus musculus

<400> 146
gactccgggc cgcgccggca gcaggagcgg aacgcagcgc agcggcggca gctgccagga
60

gatgcgagca tagaccggac tgtgagcacc tttccctctt cgggctgtaa gggagcgaga
120

cagccaccgg agcgaggcca ctccagagcc ggcagcggca ggacccggga cacaagagta
180

gccccgagac acccccagac gtagcgggcg ctccaggtga tcgagtccac gccgctccct
240

gcaggcgaca ggcgacgccc ccgcgcagcc cggccactgg ctcttccctc ccgggacaaa
300

cttttctgca agcccttgga ccaaacttgt cgcgcgtcac cgtcgcccag ccgggtccgc
360

gtagagcgct catctttagc gagatgtctg agcgcaaaga aggcagaggc aaggggaagg
420

gcaagaagaa ggaccgggga tcccgcggga agcccgcgcc cgccgaaggc gacccgagcc
480

cagcattgcc tcccagattg aaagagatga aaagccagga gtcagctgca ggctccaagc
540

tcgtgcttcg gtgtgaaacc agctctgagt actcctcact cagattcaaa tggttcaaga
600

acgggaatga gctgaaccgt aggaataaac cacaaaacgt caagatacag aagaagccag
660

ggaagtcaga gcttcgaatc aacaaagcgt ccctggctga ctctggagaa tatatgtgca
720

aagtgatcag caagttagga aacgacagtg cctctgccaa catcaccatt gttgagtcaa
780

acgacctcac cactggcatg tcagcctcaa ctgaaagacc ttatgtgtcc tcagagtctc
840

ccattagaat atcagtttca acagaaggcg caaatacttc ttcatccaca tctacatcca
900

cgactgggac cagccatctc ataaagtgtg cggagaagga gaaaactttc tgtgtgaatg
960

gaggcgagtg cttcatggtg aaggacctgt caaacccctc aagatacttg tgcaagtaag
1020

aaatgaattc ctctctgtgc ctcgtacctg taacagctta tcccagattg ttctgtgtcg
1080

ccatgaaccc ctggcttttt tttccttact ttgttacatc ttgttttaaa taattctcat
1140

ttatttgtgg agggtttttt gaaatatttg caccatctgc cattgcctct gtcatgttca
1200

gaattgattt tacttttcaa ggttttaggg tgtttttggt tcttgatggg ttgagtattt
1260

tttttgtttg gttggttttg ggtttttgct gttttgtttt gtttttttgtt tttgttttct
1320

tttttgcctt catatatata attttgcttt cctcctggtg ttccttaata gctactgaaa
1380

gaagtgtgca aatattgtag aaagctgtca ctttgaatcc ctactttttt atcccatgta
1440

ttaattgagc cataaggtac ataaggtaac tttttttaa cctcagtgct tacctgcaag
1500

gtgaacagga caaatagagg ttgcaagaga gcagaaagtt acctgctaaa gcatttctta
1560

tgctctggat tatggtattg ccccataatt agttttcaag acaaatttta agttgccctt
1620

tctagttact
1630

<210> 147
<211> 366
<212> DNA
<213> Mus musculus

<400> 147
ttcaaggcac tgctcgtcct tgctcgcact catttgccct tggatcatag gcgatggccc
60

cagctcctag cctcctgcac taccccataa tcgtctgtca cccttttgtt ttttgcagag
120

ctcacaactg gcatgtcagc ctcaactgaa agaccctatg tgtcctcaga gtctcccatt
180

agaatatcag tttcaacaga aggcgcaaat acttcttcat ccacatctac atccacgact
240

gggacaagcc atctaataaa gtgtgcggag aaggagaaaa ctttctgtgt gaacggaggc
300

gagtgcttca tggtgaagga cctgtcaaac ccctcaagat acttgtgcaa gtaagaaatg
360

aattcc
366

<210> 148
<211> 412
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (339)..(339)
<223> n = undefined nucleitide

<400> 148
cacccggccc aagttgaaga agatgaagag ccagacggga caggtgggtg agaagcaatc
60

gctgaagtgt gaggcagcag cgataaatcc ccagccttcc taccgttggt tcaaggatgg
120

caaggagctc aaccgcagcc gagacattcg catcaaatat ggcaacggca gaaagaactc
180

acgactacag ttcaacaagg tgaaggtgga ggacgctggg gagtatgtct gcgaggccga
240

gaacatcctg ggaaggaca ccgtacgagg ccggctttac gtcaacagcg tgacgaccac
300

cctgtcatcc tggtcggggc acgccgggaa gtgcaacgng acagccaagt cctattgcgt
360

caatggaggc gtctgctact acatcgaggg catcaaccag ctctcctgca ag
412

<210> 149
<211> 350
<212> DNA

<213> Homo sapiens

<400> 149
ggtcatcttc cagttttgac gtggggcatg aaggagatga ttcctggggc ctagggatag
60

tctcagtgcg tcactggcac atgtctctca taccctcagt gagcaccacc ctgtcatcct
120

ggtcggggca cgccccggaag tgcaacgaga cagccaagtc ctattgcgtc aatggaggcg
180

tctgctacta catcgagggc atcaaccagc tctcctgcaa gtaagtgacc agtaggggtg
240

ggcatgggag caagaacagg gtaggagatg ctgggtcaga agtggagggc tctaggaaaa
300

gagggttcca agccactgac aagaggtccc caaggggtgt agacaggaag
350

<210> 150
<211> 629
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (554)..(554)
<223> n = undefined nucleitide

<220>
<221> misc_feature
<222> (577)..(577)
<223> n = undefined nucleitide

<220>
<221> misc_feature
<222> (594)..(594)
<223> n = undefined nucleitide

<400> 150
gggagtcaag agatggcagt acttggctga aggttggtag tgagagatca atataatcat
60

ctggtattat tttccttctg cctggaggac ttgctttaac atttcaagta gtgtgggtct
120

gctgctgacg aattcataca aattttatac gacgacatat tccacagagc gatccgagca
180

cttcaaaccc tgccgagaca aggaccttgc atactgtctc aatgatggcg agtgctttgt
240

gatcgaaacc ctgaccggat cccataaaca ctgtcggtaa gccactgagg ccactgatgg
300

aaagggcagg cccgttgcaa ggcgtggggg tggagggtgc tggcagcatc tggtatgtgt
360

catatccggg atacacacag tcccaccgtt tgaatagcag aattgcgagt cttaatttgg
420

aaagggcaag gctgctgcct ctttaacagt ggaagaagac aaaatggaaa caaagtagtt
480

acggtttaag ttttacctga ccaagcaaac aaagatttac ttttagatct gcaaagttaa
540

tggaaataat tatntacaca ctttagaagc gtctgtntat gatgtggagc ttangcatat
600

atcctagtac tcagaaataa tctgttctt
629

<210> 151
<211> 595
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (205)..(205)
<223> n = undefined nucleitide

<400> 151
gtgtctgcgg tattcaaaaa cttttgaaac actgcatgtc caacaaaatt tatttttgt
60

gtgaatgtaa gtttttattg agggtactgt ttttcaaccc tactctcttg accaagaatg
120

aaactattta caaattaaga tgccaacaga tcacgaagag ccctgtggtc ccagtcacaa
180

gtcgttttgc ctgaatgggg ggctntgtta tgtgatacct actattccca gcccattttg
240

taggaagtga actgatgctg gcttctcttt gtcttattcc aagttgggca tgagattttc
300

cctgcattag aaggttgttg agacctgaag cctgggaagg tgcgttgaaa actatacagg
360

agctcgttgt gaagaggttt ttctcccagg ctccagcatc caaactaaaa gtaacctgtt
420

tgaagctttt gtggcattgg cggtcctagt aacacttatc attggagcct tctacttcct
480

ttgcaggaaa ggccactttc agagagccag ttcagtccag tatgatatca acctggtaga
540

gacgagcagt accagtgccc accacagtca tgaacaacac tgaagaaacg tcaaa
595

<210> 152
<211> 545
<212> DNA
<213> Homo sapiens

<400> 152
taagaaataa aggattagat ttttaattct tttacctagt ggtgtttcat tttctgcctt
60

tgtaaaataa aaacaatgat ttggttcact ttgacgtttc ttcagtgttg ttcatgactg
120

tggtgggcac tggtactgct cgtctctacc aggttgatat catactggac tgaactggct
180

ctctgaaagt ggcctttcct gcaaaggaag tagaaggctc caatgataag tgttactagg
240

accgccaatg ccacaaaagc ttcaaacagg ttacttttag tttggatgct ggagcctggg
300

agaaaaacct cttcacaacg agctcctgta tagttttcaa cgcaccttcc caggcttcag
360

gtctcaacaa ccttctaatg cagggaaaat ctcatgccca acttggaata agacaaagag
420

aagccagcat cagttcactt cctacaaaat gggctgggaa tagtaggtat cacataacaa
480

agccccccat tcaggcaaaa cgacttgtga ctgggaccac agggctcttc gtgatctgtt
540

ggcat
545

<210> 153
<211> 715
<212> DNA
<213> Homo sapiens

<400> 153
gcctgagctg ggcagggggc ggaggcgggg gctcggctgt ctccggggct gccacgcaga
60

gcgggcttcg tggcgtggat gaagaaactg aggcacagag ggattaagta gcctgctcaa
120

gatcacacag ctagtaagga accaagattc aaacttgggc agtgtgattc agagacttta
180

aattcaacgc tggtgcctca ctgcctcaca ctaaaagtga atcagaaaaa taaagaacca
240

gcatcaaatt tgaagtggcc acaaattcta ttaaagcaga agaaatagtg gtgaaccata
300

aaagataacc agtttcctct ctattctgca atttagagga aaaattttca tccaaggaca
360

gatcaggtgg tggacctaga tgggaaaccc aaattataat caagagattt cttggtactg
420

tttttcaacc ctactctctt gaccaagaat gaaactattt acaaattaag atgccaacag
480

atcacgaaga gccctgtggt cccagtcaca agtcgttttg cctgaatggg gggctttgtt
540

atgtgatacc tactattccc agcccatttt gtaggaagtg aactgatgct ggcttctctt
600

tgtcttattc caagttgggc atgagatttt ccctgcatta gaaggttgtt gagacctgaa
660

gcctgggaag gtgcgttgaa aactatacag gagctcgttg tgaagaggtt tttct
715

<210> 154
<211> 669
<212> DNA
<213> Mus musculus

<400> 154
gagtgttcaa acacttgtga aacgctgcat gtctagcaaa attttctttt tttatgggaa
60

tataaatttc tgttgaggtg ctgattttca accttaattc ttccatcaag aatgaaacta
120

tttaaaaatt aagatgccaa caggtaattt cttatcacga gcagccctgt ggtcccaggc
180

acaggtcatt ttgcctcaat ggggggattt gttatgtgat ccctactatc cccagcccat
240

tctgtaggaa gtgaactgtt gctggcttct ctttgtctta ttccaagttg ggtcatgaga
300

ttttccctgc accctgggaa ggtgcattga aaattacacc ggagcacgct gcgaagaggt
360

ttttctccca agctccagca tcccaagcga aagtaatctg tcggcagctt tcgtggtgct
420

ggcggtcctc ctcactctta ccatcgcggc gctctgcttc ctgtgcaggg ccgagtggaa
480

ctgaccctcc aggacatatg tgagatgcta aaaggaagac taaagaagtg gaagggccac
540

cttcagaggg ccagttcagt ccaatgtgag atcagcctgg tggaaacaaa caataccaga
600

acccgtcaca gccacagaaa acactggaaa catacatccc cagggaaggg catcattacc
660

tacaaaggg
669

<210> 155
<211> 614
<212> DNA
<213> Mus musculus

<400> 155
gagtgttcaa acacttgtga aacgctgcat gtctagcaaa attttctttt tttatgggaa
60

tataaatttc tgttgaggtg ctgattttca accttaattc ttccatcaag aatgaaacta
120

tttaaaaatt aagatgccaa caggtaattt cttatcacga gcagccctgt ggtcccaggc
180

acaggtcatt ttgcctcaat ggggggattt gttatgtgat ccctactatc cccagcccat
240

tctgtaggaa gtgaactgtt gctggcttct ctttgtctta ttccaagttg ggtcatgaga
300

ttttccctgc accctgggaa ggtgcattga aaattacacc ggagcacgct gcgaagaggt
360

ttttctccca agctccagca tcccaagcga aagtaatctg tcggcagctt tcgtggtgct
420

ggcggtcctc ctcactctta ccatcgcggc gctctgcttc ctgtgcagga agggccacct
480

tcagagggcc agttcagtcc agtgtgagat cagcctggta gagacaaaca ataccagaac
540

ccgtcacagc cacagagaac actgaagaca tacatcccca gtgaagggca tcattaccta
600

caaaggcgga ctgg
614

<210> 156
<211> 513
<212> DNA
<213> Homo sapiens

<400> 156
ttaagaaata aaggattaga tttttaattc ttttacctag tggtgtttca ttttctgcct
60

ttgtaaaata aaaacaatga tttggttcac tttgacgttt cttcagtgtt gttcatgact
120

gtggtgggca ctggtactgc tcgtctctac caggttgata tcatactgga ctgaactggc
180

tctctgaaag tggcctttcc tgcaaaggaa gtagaaggct ccaatgataa gtgttactag
240

gaccgcccat gccacaaaag cttcaaacag gttactttta gtttggatgc tggagcctgg
300

gagaaaaacc tcttcacaac gagctcctgt atagttttca acgcaccttc ccaggcttca
360

ggtctcaaca accttctaat gcagggaaaa tctcatgccc aacttggaat aagacaaaga
420

gaagccagca tcagttcact tcctacaaaa tgggctggga atagtaggta tcacataaca
480

aagcccccca ttcaggcaaa acgacttgtg act
513

<210> 157
<211> 243
<212> DNA
<213> Sus scrofa

<400> 157
aagagccctg tggtcccagt cacaggtcat tttgcctgaa tggagggatt tgttatgtga
60

tacctactat tcccagcccc ttttgtagga agtgaactga tgctggcttc tctttgtctt

120

attccaagtt ggggcatgag attttgcctg cattagaagg ttgttgagac ctgaagcctg
180

gtaaggtcat gcagaacatt gaagaaatac catagtgaac tcaaaatcgt tgcttctttg
240

tta
243

<210> 158
<211> 300
<212> DNA
<213> Sus scrofa

<220>
<221> misc_feature
<222> (111)..(275)
<223> n = undefined nucleitide

<400> 158
aagagccctg tggtcccagt cacaggtcat tttgcctgaa tggagggatt tgttatgtga
60

tacctactat tcccagcccc ttttgtagga agtgaactga tgctggcttt ncnttggcct
120

aatnccagnt tgggcatgag atttgcctgc attagaangg tgttgaganc tgaagcctgg
180

taaaggcatg cagaacattg aagaatacnt agtgaactcc aaatcggtgc ttccttggta
240

caaaaggcgn aatgnagccc atacggtaaa gatcnatgag ttaatcctcc ttggccccaa
300

<210> 159
<211> 2360
<212> DNA
<213> Mus musculus

<400> 159
ttgtttgttg ttgcatacac caggctgctg gacactgaac ttctggcaat tctcttgtct
60

ctgaccccat ctcctggtag aggtgcactg gactacagac atgtgcccta ctgcactggc
120

tatttatgtg gatttgaact caggtcatca ggctgtgggg cgagtgcctt accctctgaa
180

ctatcttccc agcccctgtt gttggcttgt gtctcatgtg ttagggaggt tcagtgccct
240

catggcactt ggcagtgctt tgtgaggcac cagagagttg gaggccacca tggtgtgaca
300

tgaccctttg catgtccttc cagctatttc tcaggctgga tacaaagtgc caggtgcatg
360

gaaacttcat tatagaggtt caggtaccca ggtcaatgtt ttcctcagga actctaagta
420

gaaaactaaa ctctagtcag tttgctatta aaaacagatc ccagctcaag cgtcccggga
480

ctccttttgt accctggaca tctggttgac agttctcatc cttcaacttg ctcagccctc
540

tgggtctcag atcagtagcc agccacatag aagcaaacac tcttttaatc gggtacttgg
600

ccaccccctt cctcccctaa gacgagggga atactcacac acatgctggc ttctcttcct
660

gcaccaaaaa ccggcaggtt ccatggaagc agtactgagt gtgggaatct gggcacttgt
720

tgaagtgaga caccactgca gccgccacgg gtgagtctgc tggggcaaag agacatcatc
780

agacctggca cagctcacac ccaggaggaa tttctgccct cacctgatgc cttctgcaaa
840

actcacgtcc taatgcccag ccagggctca gagttttcat taagcagtct gtatattttt
900

ctaagataac aaaataattt ctccaaaggc tttggtataa ttcaaagata gctagttaga
960

ttcatttgca aaatggcaca cacctgaaat cccagcactc agaaggtaga ggcacaagga
1020

tcaggagttc gaggccaacc tagtccatat gtggagtttg aggtcagttg gatctcatta
1080

tctccaaccc ccaaaagaag ccaaggaacg gctcagtagg caaagtgctt gctgtgcaaa
1140

gatgaggacc tgagcttgga ctccagcacc cacataaaga gacatcacag taaggattgc
1200

aactccagca ttctagttcc tggggaacca ctatactgct gaaggcagag ctctatgcct
1260

tgtaacagaa taaacaaaga tgctcaatgg ataaacatac tgacacacat gtaggatgga
1320

ctcaacattc tgtgttcaga gtctttgaag gagtcattgt aagcaaaggc agaaacctcc
1380

tcaatgacat cccaaagatt cctgccagtg ccccttctcc tgtgtcatca tacagcccaa
1440

aagctggggt ccacaccatg aagaaactcc acatgacacc caaaggtttg tctctctgtc
1500

cctggagcat agggtgagaa tgagaagcct gctacttctg attctctggt ttctgagcct
1560

caagtagttc aggctggcct agaattcact gtgtagccca ggctagtctt gaactcttga
1620

tcctcctgcc tccaccccca ccaagtgttg gggttatagg agtgtgatgc cactcctggt
1680

ttattcagta ttgggattga aaccaggcca gcactctaca actctacctc atcccagccc
1740

acttctggtc cttcatacag ccaactatct tcctgctact tataataaat gcttccagtc
1800

ctttctgctg cccttctcca ggctaagaga agaaaggatg aaggaagagg aatgacaatc
1860

catgctatga caactaaatg gtagctaaaa ataaaacaac cctttgcttt aattacagtg
1920

atacatacac ttttgaaact tttccagaag cttttctgaa tggcaaaggt agttcactga
1980

aactactgac atagaataaa atccacctta gagaataaag cacatcttaa tcctcaactc
2040

atcaagagtc ataaaaacac agcacacacc aatgacatac ttgtgaactt acattcctgt
2100

tctaaaaatc aagggtgaat cacattgcaa ccaggaaact gcccttgcct gggactcagg
2160

ggcagctgcc aaagcacaga actggtaagt ttacgaggag actccaagtt cccgatatct
2220

tcccccaaga ttggaccttt caactctttt tctcttttta ttcttttaaa ttaaaagatg
2280

tgtgcgttgt gtgtgtgtgt gcgcacgcgc ttgtgactgc aaatgctgcc aagtgaactt
2340

ggacaagcat tactgcatct
2360

<210> 160
<211> 180
<212> DNA
<213> Homo sapiens

<400> 160
gatctgagcc ctgcatcttt cctctcccca gcagacccgc ccgtggctgc agcagtggtg
60

tcccatttta atgactgccc agattcccac actcagttct gcttccatgg aacctgcagg
120

tttttggtgc aggaggacaa gccagcatgt gtgtaagtat cccctgttct cctggagatc
180

<210> 161
<211> 129
<212> DNA
<213> Homo sapiens

<400> 161
cagttcagac agaagacaat ccacgtgtgg ctcaagtgtc aataacaaag tgtagctctg
60

acatgaatgg ctattgtttg catggacagt gcatctatct ggtggacatg agtcaaaact
120

actgcaggt
129

<210> 162
<211> 120
<212> DNA
<213> Homo sapiens

<400> 162
cagacagaag acaatccacg tgtggctcaa gtgtcaataa caaagtgtag ctctgacatg
60

aatggctatt gtttgcatgg acagtgcatc tatctggtgg acatgagtca aaactactgc
120

<210> 163
<211> 129
<212> DNA
<213> Mus musculus

<400> 163
tagttcagat ggaagacgat ccccgtgtgg ctcaagtgca gattacaaag tgtagttctg
60

acatggacgg ctactgcttg catggccagt gcatctacct ggtggacatg agagagaaat
120

tctgcagat
129

<210> 164
<211> 1299
<212> DNA
<213> Homo sapiens

<400> 164
gacacagcca acgtggggtc ccttctaggc tgacagccgc tctccagcca ctgccgcgag
60

cccgtctgct cccgccctgc ccgtgcactc tccgcagccg ccctccgcca agccccagcg
120

cccgctccca tcgccgatga ccgcggggag gaggatggag atgctctgtg ccggcagggt
180

ccctgcgctg ctgctctgcc tgggtttcca tcttctacag gcagtcctca gtacaactgt
240

gattccatca tgtatcccag gagagtccag tgataactgc acagctttag ttcagacaga
300

agacaatcca cgtgtggctc aagtgtcaat aacaaagtgt agctctgaca tgaatggcta
360

ttgtttgcat ggacagtgca tctatctggt ggacatgagt caaaactact gcaggtaata
420

tgtcagaaat aaacaaacac agtttgtaaa attttgtttt atagatttag gggtacaagt
480

gcagatttgc tagtggatat attcagtagt ggtgaagtct gagcttttag agtacctacc
540

cctcaaatag tgtgcatgga acccattagg taatttttca tcccttaacc cccccaaaac
600

tcttctacct tttgaagtct ccagagtcta ttactccact ctctatgaca atgtgtacac
660

attatttagc tcccacttgt gagaacatgt gataaacaaa tgcagtttta ctctttgtat
720

ttctattttt ataatttgaa attaccctat atttccatgg gctgttaaat gcagtatata
780

tattattaga aacttttctg agtttttaaa aattaggtag taaatagtag cttttaaatt
840

gcacacatat gtcagaggtg cagagcaggg aggacttctg atgcttctca cacttgccaa
900

gatggtgtct ctctgctttg gatcttttcc ttcaatttct atatcaggta ttgttttaag
960

aattgattcc aggccggacg cgttggctca tgcctgtaat cccagcactt tgggaggccg
1020

aggcgggcgg atcacggggt caggagatca agaccatcct ggcgaacacg gtgaaacccc
1080

gtctctacta aaaatacaaa aaaaaaaaaa attagccagg ggtagtggcg gacgcctgaa
1140

gtcccagcta ctcgggaggc tgaggcagga gaatggcatg aacccggggg gtggagcttg
1200

cagtgagcgg agatcatgcc actgtactcc agcctgggca acacagcgag actccgtctc
1260

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaa
1299

<210> 165
<211> 1215
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (554)..(839)
<223> n = undefined nucleitide

<400> 165
taatacgaag acacagccaa cgtggggtcc tttctcggct gacagccgct ctccagccac
60

tgccgcgagc ccgtctgctc ccgccctgcc cgtgcactct ccgcagccgc cctccgccaa
120

gccccagcgc ccgctcccat cgccgatgac cgcggggagg aggatggaga tgctctgtgc
180

cggcagggtc cctgcgctgc tgctctgcct gggtttccat cttctacagg cagtcctcag
240

tacaactgtg attccatcat gtatcccagg agagtccagt gataactgca cagctttagt
300

tcagacagaa gacaatccac gtgtggctca agtgtcaata acaaagtgta gctctgacat
360

gaatggctat tgtttgcatg gacagtgcat ctatctggtg gacatgagtc aaaactactg
420

caggtaatat gtcagaaata aacaaacaca gtttgtaaaa ttttgtttta tagatttagg
480

ggtacaagtg cagatttgct agtggatata ttcagtagtg gtgaagactg ctattactcc
540

atgtgcttcc cgcnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
600

nnnnnnnnnn nnnnnggnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
660

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
720

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
780

nnnnnccnnn nnnnnnnncn nngnnnnngn nnnnnngnnn nnnnnnnnnn gttnttttng
840

aaactttttt tttgaggttt ttaaaaaaat taggggtagt aaaaataggg aggtttttta
900

aaatttgccc caccattatg tccaaaagtg gccacaagtc aggaaaggaa ccttttggag
960

ggcttttctc cccccttttgc ccccggaagg ggggtcctcc tccgggcctt gggaatcttt
1020

tttcccttac attttccaaa attccgggga ttttgttttt taaaaaaatg gagatttccc
1080

cgcgccccgg acgccgtatg ggcttcatg gccctggaaa ccccacccca ctcttttgtg
1140

gggggtcccg aggcaggggg gggggaattc cgcgggggcc ccggggaaat tacaaacacc
1200

ctccccctgg ggcga
1215

<210> 166
<211> 549
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (355)..(355)
<223> n = undefined nucleitide

<400> 166
atcccgggga gaaagccacc cggcccaagt tgaagaagat gaagagccag acgggacagg
60

tgggtgagaa gcaatcgctg aagtgtgagg cagcagcggg taatccccag ccttcctacc
120

gttggttcaa ggatggcaag gagctcaacc gcagccgaga cattcgcatc aaatatggca

180

acggcagaaa gaactcacga ctacagttca acaaggtgaa ggtggaggac gctggggagt
240

atgtctgcga ggccgagaac atcctgggga aggacaccgt cggaggccgg ctttacgtca
300

acagcgtgac gaccaccctg tcatcctggt cggggcacgc ccggaagtgc aacgngacag
360

ccaagtccta ttgcgtcaat ggaggcgtct gctactacat cgagggcatc aaccagctct
420

cctgcaaggc acctgggctg cactgcttag aacttggtac ccagagccac cacttcccca
480

tctcagcctc ccctggttcc agccaaggtt cctggaacca acttccccaa caccctttgt
540

cagccctcg
549

<210> 167
<211> 362
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (323)..(323)
<223> n = undefined nucleitide

<400> 167
agcacagctc tgaggacctg gtgttctgac cgcatctcca ccagggctgc cctctccccc
60

gagggctgac aaagggtgtt ggggaagttg gttccaggaa ccttggctgg aaccagggga
120

ggctgagatg gggaagtggt ggctctgggt accaagttct aagcagtgca gcccatgtgc
180

cttgcaggag agctggttga tgccctcgat gtagtagcag acgcctccat tgacgcaata
240

ggacttggct gtctcgttgc acttccgggc gtgccccgac caggatgaca gggtggtgct
300

cacgctgttg acgtaaagcc ggncccggac ggtgtccttc cccaggatgt tctcggcctc
360

gc
362

<210> 168
<211> 458
<212> DNA
<213> Mus musculus

<400> 168
gtgtgaggca gcggcgggaa acccccagcc ctcctatcgc tggttcaagg atggcaagga
60

actcaaccgg agtcgtgata ttcgcatcaa gtatggcaat ggcagtgagc accactctgt
120

catcctggtc gggacatgcc cggaagtgca atgagaccgc caagtcctac tgtgtgaatg
180

gaggcgtgtg ctactacatc gagggcatca accagctctc ctgcaaaggc tgaggagctg
240

taccagaaga gagtgctgac aattactggt atctgtgtgg ccctgctggt cgtgggcatc
300

gtctgtgtgg tcgcctactg caagaccaaa aaacagagga ggcagatgca tcatcatctc
360

cggcagaaca tgtgcccagc ccaccagaac cgaagcctgg ccaacgggcc agccaccctc
420

ggctggacca tgaggagacc agatggcaga ttaatctc
458

<210> 169
<211> 539
<212> DNA
<213> Danio rerio

<400> 169
ccaccagcag agccacgcag atgccagtta tcgtcagcac tcgttttggt acagctcctc
60

agccttgtag aaaccggcca taacggaggt ttgacagcgt tcgccggtat agtcatttgg
120

acacttgcag gacagctgat ttataccatg tatgaaataa cagtctccac cgttgatgca
180

gtatgtcttc tcagtttcat tgcacttcct ggcatgactt gagcccggag acaatgtggt
240

ggttatgctt tggacgctga cgaagctggt ggcgttttct ctgcccagcg agttctccac
300

cacacaggtg tagttcccag aatcctccag tctgactttg ctaatgtgaa gctttgagtt
360

tttcttgttg gttttgattt tgacggtttt cttttggcga agctggctgc catctttgta
420

ccagttgaag gaggggctcg ggttgcccac agcttcacac ttcagtgtca actttttacc
480

ttcctggagc cactgagaat ccatgggctt cacctttgga gctgatgcgc agtctttac
539

<210> 170
<211> 654
<212> DNA
<213> Gallus gallus

<400> 170
cacgctggga gatgagtgct gtggtgccca gctgtgaggt gcctgggctg gcagtgcttc
60

tccctctctc cctctgcagg ggaaagaaag aagggacttt ttctttctct gaagtagaag
120

ttcagatttt gatggtaagg gagctgatgt ggaggcctgg ccttaaggaa ggctttcagt
180

aggcagtaca gtctttggag ctgctgcagc agacctggcg gttgtctacc ttgcaatttg
240

agtatgacag aagagtagcc tgtggattcc actatactac aacgtattcc actgagcgat
300

ctgagcactt taagccatgc aaagacaagg atcttgcata ctgtctcaac gagggggaat
360

gctttgtgat tgaaacctta acaggatcac ataaacactg ccgcagcaat tgcccttctg
420

gtgttttctg ctggtgacct gtctgaatag atgttcttcc agaggtggtt gtggtttggg
480

gcattgatgc tgggaagagg attaccagga agagctcagc tgttccttca ttgctcagtc
540

cacgtttata aagaaggatg gacagtgacc tgtgagcaag cttgtttgca aaagaaagca
600

ttatctgttg gtaacttttg caataaaaaa tatttcttgt attactctaa aaaa
654

<210> 171
<211> 758
<212> DNA
<213> Gallus gallus

<220>
<221> misc_feature
<222> (4)..(4)
<223> n = undefined nucleitide

<400> 171
gcanggcggg aggcgccgcg cggtcgctgt ccgcgggcag acagcggcat tacataaccg
60

cgtacagaga gcagctgcgg gattacacga tgcagattag cggcggcgtt gattcagcag
120

atgccctgtg cgtgtgtgag gggattacg gcggcgcggg gcagaaccgc cgtgcgggtg
180

ccgttttaga agaatagctt ctgaccaaga attagaattg ttggaataat atgcgaacag
240

atcatgaaga actctgtggc accagttatg gatctttttg tctaaatgga ggcatttgct
300

atatgattcc tactgtaccc agtccattct gcagacatct tccgaaagca gcaaaccaag
360

cttcagcctt acataagtca gtcttctcta tcttcgtttt acatacagac accactgcac
420

tcccaagctg ccatttaatg cctgctcatt tctatacgca atgaaagata actagaaaat
480

ccgtatttca aggctatcct ccatttctac atccctgcaa actacctaag aacaattaga
540

tggaacagga ttgtctacaa cattgttatc acaaaggagg ctatcttatg gatggaattt
600

ctttttctc agatgtatta cttaccagca aggaaggtag ttctgtttga atcttctcaa
660

taaacaccac atttcctgtt tcaggttggg tgggaactat tcttcaaacg gaggaggttt
720

atgtgttcct ttcgttccta taatgtctca ataatgag
758

<210> 172
<211> 547
<212> DNA
<213> Mus musculus

<400> 172
gttgctgaag tcctcagtgt tcaaacactt gtgaaacgct gcatgtctag caaaattttc
60

tttttttatg ggaatataaa tttctgttga ggtgctgatt ttcaacctta attcttccat
120

caagaatgaa actatttaaa aattaagatg ccaacagatc acgagcagcc ctgtggtccc
180

aggcacaggt cattttgcct caatgggggg atttgtattg atccctacta tccccaccca
240

ttctgtaggt tttatcattt gtttctaaga cattgcctac ttaaaccatt cgtgcaattg
300

ggcaccttgg tgtacccagt gtttctgaag gagttattcc attgacgcgc cccaagttct
360

tcatgcagtg gtgttcctga atgcttgaaa tctgttttct gcgaatcctt ggtgggatgg
420

ctagaaacct gtgaaaaatc atgaaatcac caaataccat gtgatgtgta tagtctcttc
480

tcctctccac tgacagctta atcaggggaa agggactgtt gctgcttctc tttgtcttat
540

tcccagt
547

<210> 173
<211> 233
<212> DNA
<213> Homo sapiens

<400> 173
cggatgtatc ccaacaccgt cacggaaata ttctgctgac attgcatgtt actgcttcca
60

ggtgctctat atatttgcat tctccgtgaa tgcagaaatt ttgaaattct gcatcacatg
120

gattttcttt ctttctgttt cttctatttt ttccattttt gcctcccttt ttctttcttt
180

tgggtttatc tgaagtattt tcactttccg gcttgtgttg ggcgataaca tca
233

<210> 174
<211> 533
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (7)..(7)
<223> n = undefined nucleitide

<400> 174
ccctagntgc caccacacaa tcaaagtgga aaggccactc ctctaggtgc cccaagcaat
60

acaagcatta ctgcatcaaa gggagatgcc gcttcgtggt ggccgagcag acgccctcct
120

gtgtccctct ccggaaacgt cgtaaaagaa agaagaaaga agaagaaatg gaaactctgg
180

gtaaagatat gactcctatc aatgaagata ttgaagagac aaatattgct tataaggcta
240

tgaagttacc tccaggttgg tggcaagctg caaagtgcct tgctcatttg aaaatggaca
300

gaatgcgtct caggaaaaca gctagtagac atgaatttta aataatgtat ttacttttta
360

tttgcaactt cagtttgtgt tattattttt taataagaac attaattata tgtatattgt
420

ctagtaattg ggaaaaaagc aactggttag gtagcaacaa cagaagggaa atttcaataa
480

cctttcactt aagtattgtc accaggatta ctagtcaaac aaaaaaaaaa aaa
533

<210> 175
<211> 689
<212> DNA
<213> Mus musculus

<220>
<221> misc_feature
<222> (671)..(671)
<223> n = undefined nucleotide

<400> 175
gcagattatt tgtttaccac ttagaacaca ggatgtcagc gccatcttgt aacgacgaat
60

gtgggggcgg ctcccaacac ttcaccatgg ttttgacctt gtcatgacca gttattttct
120

ggcttatctc cactaatctt gggagcctca gcaccagccc tgagttcata tcacaccacc

180

aaagtctttg acctggaaga gctttaactt cctaagcctc ctgcttccac tgggcagcac
240

tggtacccgg agaatcctgt gtcccttgtc tactccatcc tgttctgcag gtcttgcaat
300

tctccactgt gtggtagcag atgggaacac aaccagaaca ccagaaacca atggctctct
360

ttgtggagct cctggggaaa actgcacagg taccacccct agacagaaag tgaaacccca
420

cttctctcgg tgccccaagc agtacaagca ttactgcatc catgggagat gccgcttcgt
480

ggtggacgag caaactccct cctgcatggc ccggctcagc atctacttgt ggagaaactg
540

acgcagactt tcctcctgaa atctgaatat gagaaaccag gtccagttct gccctgctgg
600

tgtcccaact cccttgtgca agaaaaggcg attctaatcg tgttaggatg ctcgatagtt
660

ccaatcatct nctgggtgtt tcaatgaaa
689

<210> 176
<211> 1196
<212> DNA
<213> Cercopithecus aethiops (African green monkey)

<400> 176
gcccagcgga atctcttgag tcccaccgcc cagctccggt gccagcgccc agtggccgcc
60

gcttcgaaag tgactggtgc ctcgccgcct cctctcggtg cgggaccatg aagctgctgc
120

cgtcggtggt gctgaagctc cttctggctg cagttctttc ggcactggtg actggcgaga
180

gcctggagca gcttcggaga gggccagctg ctggaaccag caacccggac ccttccactg
240

gatctacgga ccagctgcta cgcctaggag gcggccggga ccggaaagtc cgtgacttgc
300

aagaggcaga tctggacctt ttgagagtca ctttatcctc caagccacaa gcactggcca
360

caccaagcaa ggaggagcac gggaaaagaa agaagaaagg caagggacta gggaagaaga
420

gggacccatg tcttcggaaa tacaaggact tctgcatcca cggagaatgc aaatatgtga
480

aggagctccg ggctccctcc tgcatggcag ctgggcagaa agatgttact tgatttgttt
540

ggtttgtcct gtgatgaaag aggcctggta gctcagcgtt cagaggccaa aggccagagc
600

tgccacccag gttaccatgg agagaggtgt catgggctga gcctcccagt ggaaaatcgc
660

ttatatacct atgaccatac aactatcctg gctgtggtgg ccgtggtgct gtcctctgtc
720

tgtctgctgg tcatcgtggg gcttctcatg tttaggtacc ataggagagg tggttatgat
780

gtggaaaacg aagagaaagt gaagttgggc atgactaatt cccactgaga gacttgtgct
840

caaggaatca gctggtgact gctacctctg agaagacaca aggtgatttc agattgcaga
900

ggggaaagac gtcacatcta gccacaaaga ctccttcatc cccagtcgcc atctaggatt
960

gggcctccca taattgcttt gccaaaatac cagagccttc aagtgccaaa ccgagtatgt
1020

ctgatagtat ctgggtgaga agaaagcaaa agcaagggac cttcatgccc ttctgattcc
1080

cctccaccaa gccccacttc cccttataag tttgtttaag cactcacttc tggattagaa
1140

tgccggttaa attccatatg ctccaggatc tttgactgaa aaaaaaaaaa aaaaaa
1196

<210> 177
<211> 564
<212> DNA
<213> Homo sapiens

<400> 177
acggggtccg agaaagttaa gcaactacag gaaatggctt tgggagttcc aatatcagtc
60

tatcttttat tcaacgcaat gacagcactg accgaagagg cagccgtgac tgtaacacct
120

ccaatcacag cccagcaagc tgacaacata gaaggaccca tagccttgaa gttctcacac
180

ctttgcctgg aagatcataa cagttactgc atcaacggtg cttgtgcatt ccaccatgag
240

ctagagaaag ccatctgcag gtgtctaaaa ttgaaatcgc cttacaatgt ctgttctgga
300

gaaagacgac cactgtgagg cctttgtgaa gaattttcat caaggcatct gtagagatca
360

agtgagccca aaattaaagt tttcagatga aacaacaaaa cttgtcaagc tgactagact
420

cgaaaatatg gaaagttggg gatcacaatg aaatgagaag ataaaatcag cggtggccct
480

tagactttgc catccttaag gagtgatgga agccaagtga acaagcctca gtgacacaag
540

tcaaattcat aggttcactc tggg
564

<210> 178
<211> 387
<212> DNA
<213> Homo sapiens

<400> 178
ggcacgaggg aggctctttg ttatagatgc ttttgccccc ttaatacagc aatgagagca
60

ctgaccgaag aggcagccgt gactgtaaca cctccaatca cagcccagca agctgacaac
120

atagaaggac ccatagcctt gaagttctca cacctttgcc tggaagatca taacagttac
180

tgcatcaacg gtgcttgtgc attccaccat gagctagaga aagccatctg caggtgtcta
240

aaattgaaat cgccttacaa tgtctgttct ggagaaagac gaccactgtg aggcctttgt
300

gaagaatttt catcaaggca tctgtagaga tcagtgagcc caaaattaaa gttttcagat
360

gaaacaacaa aacttgtcaa gctgact
387

<210> 179
<211> 389
<212> DNA
<213> Homo sapiens

<400> 179
ggcacgagga aagttaagca tctacaggtt atggctttgg gagttccaat atcagtctat
60

cttttattca acgcaatgac agcactgacc gaagaggcag ccgtgactgt aacacctcca
120

atcacagccc agcaaggtaa ctggacagtt aacaaaacag aagctgacaa catagaagga
180

cccatagcct tgaagttctc acacctttgc ctggaagatc ataacagtta ctgcatcaac
240

ggtgcttgtg cattccacca tgagctagag aaagccatct gcaggtgtct aaaattgaaa
300

tcgccttaca atgtctgttc tggagaaaga cgaccactgt gaagcctttg tgaagaattt
360

tcatcaaggc atctgtagag atcagtgag
389

<210> 180
<211> 409
<212> DNA
<213> Homo sapiens

<400> 180
aactacagga aatggctttg ggagttccaa tatcagtcta tcttttattc aacgcaatga
60

cagcactgac cgaagaggca gccgtgactg taacacctcc aatcacagcc cagcaagctg
120

acaacataga aggacccata gccttgaagt tctcacacct ttgcctggaa gatcataaca
180

gttactgcat caacggtgct tgtgcattcc accatgagct agagaaagcc atctgcaggt
240

gtctaaaatt gaaatcgcct tacaatgtct gttctggaga aagacgacca ctgtgaggcc
300

tttgtgaaga attttcatca aggcatcttg tagagatcaa gtgagcccaa aattaaagtt
360

ttcagatgaa acaacaaaac ttgtcaagct gactagactc gaaaatatg
409

<210> 181
<211> 568
<212> DNA
<213> Homo sapiens

<400> 181
ccgtcagtct agaaggataa gagaaagaaa gttaagcaac tacaggaaat ggctttggga
60

gttccaatat cagtctatct tttattcaac gcaatgacag cactgaccga agaggcagcc
120

gtgactgtaa cacctccaat cacagcccag caaggtaact ggacagttaa caaaacagaa
180

gctgacaaca tagaaggacc catagccttg aagttctcac acctttgcct ggaagatcat
240

aacagttact gcatcaacgg tgcttgtgca ttccaccatg agctagagaa agccatctgc
300

aggtgtctaa aattgaaatc gccttacaat gtctgttctg gagaaagacg accactgtga
360

ggcctttgtg aagaattttc atcaaggcat ctgtagagat cagtgagccc aaaattaaag
420

ttttcagatg aaacaacaaa acttgtcaag ctgactagac tcgaaaataa tgaaagttgg
480

gatcacaatg aaatgagaag ataaaattca gcgttggcct ttagactttg ccatccttaa
540

ggagtgatgg aagccaagtg aacaagcc
568

<210> 182
<211> 282
<212> DNA
<213> Homo sapiens

<400> 182
atggctttgg gagttccaat atcagtctat cttttattca acgcaatgac agcactgacc
60

gaagaggcag ccgtgactgt aacacctcca atcacagccc agcaagctga caacatagaa
120

ggacccatag ccttgaagtt ctcacacctt tgcctggaag atcataacag ttactgcatc
180

aacggtgctt gtgcattcca ccatgagcta gagaaagcca tctgcaggtg tctaaaattg
240

aaatcgcctt acaatgtctg ttctggagaa agacgaccac tg
282